The Image Survival and Protection of Intangible Cultural Heritage in the New Media Environment

Jinxia Wang (✉ 1754354519@qq.com)
Shaanxi Fashion Engineering University

Research Article

Keywords: intangible cultural heritage, image survival, computer technology, deep learning

Posted Date: January 10th, 2023

DOI: https://doi.org/10.21203/rs.3.rs-2144055/v1

License: ☺ ☀ This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License
Abstract

As a product of agricultural civilization, intangible cultural heritage (ICH) has been in a bad situation in recent years. Modern video media, with the dual identity of art and media, is an effective way to preserve and disseminate ICH. First, a hybrid network composed of a Bi-directional Long Short-Term Memory (Bi-LSTM) network with attention structure and Neural Network is adopted to extract relevant knowledge. Then, the generative adversarial network (GAN) is optimized. Lastly, this model is tested. The test results reveal that in the dataset constructed here, when the resolution of the processed image is 48×64×48, it takes 0.4825s for the unimproved GAN to process the image, while the algorithm improved only needs 0.0391s to process the image, with a speedup of 12.2.

Introduction

ICH is also known as "spiritual cultural heritage" or "national and folk cultural heritage". According to the definition in the International Convention for the Protection of Intangible Cultural Heritage approved by the United Nations Educational, Scientific and Cultural Organization in 2003, ICH means a variety of practices, performances, expression forms, knowledge and skills, and related tools, objects, handicrafts as well as cultural sites that are taken as their cultural heritage by some groups and sometimes individuals [1, 2]. ICH is the commonwealth of mankind, the essence of ancient human culture, and the valuable wealth left over by all ethnic groups in history. However, with the advent of the information age and the acceleration of the global modernization process, China's cultural ecological environment has undergone earth-shaking changes. ICH's living environment is worse and worse, and has been impacted unprecedentedly. Although there are multiple ways to protect ICH in China, many of the ICH passed down by word of mouth among the people are dying out. Currently, ICH has several problems to be solved. First, some missing contents in the investigation need to be supplemented, especially the image records. Second, it is the preservation and protection of achievements. Most of the primary and secondary materials are still in villages and towns. How to keep and collect these original materials is an urgent problem to be solved. Third, it is inheritance, sorting, research, development and utilization. Multiple provinces and autonomous regions in China are short of talents in this field, such as anthropology, ethnology, folklore, archival science, editorial science, and proofreading. If it is preserved first, passed on, and then sorted out and studied in batches, it will be ensured that funds and talents are available for development and utilization. If people do not hurry up the preservation, this precious national culture may disappear once the inheritor of an ICH passes away [3, 4].

In relevant research fields, researchers Hu and Yang (2022) [5] pointed out that ICH protection faced a survival crisis with economic globalization. ICH is the wisdom crystallization of a country’s laboring people of all generations. It has distinctive national features. The application of digital technology to ICH protection plays a great role in promoting ICH inheritance, transformation as well as utilization. After the informatization development strategy is proposed by China, it has become more realistic to protect cultural heritage by using digital technology. The management mechanism as well as format specification are not perfect, and ICH has particularity, so developing digital means remains difficult. With
the purpose of better protecting digital technology, the vital role, in-depth development and application of
digital technology in ICH protection and inheritance were studied. Yan and Chiou (2021) [6] pointed out
that the living environment of Chinese cultural heritage was deteriorating with economic and social
development. The inheritance and protection of ICH have become a common concern of mankind. The
application of reasonable and scientific protection methods is of great significance to ICH inheritance
and protection. The path of ICH inheritance and protection was explored. Lu and Zheng (2021) [7] used
modern computer digital technology to protect ICH. Digital computer technology was introduced into
these protection methods to improve their shortcomings, and the method of using computer data to
protect ICH was introduced in detail.

To sum up, worldwide scholars have conducted in-depth research on the inheritance, protection,
development and utilization of ICH from different disciplines and perspectives. In recent years, the
academic community has reached a consensus on the participation of multiple subjects in ICH
protection. It is generally believed that the government should play a role in ICH protection, and social
forces should play a more critical role. Its related theoretical research results and practical exploration
have great reference value for this exploration. However, there are still great research deficiencies.
Thereby, this exploration restores ICH images based on image restoration technology in deep learning
(DL) technology. The hybrid network composed of a Bi-directional Long Short-Term Memory (Bi-LSTM)
network with attention structure and Neural Network (NN) for image generation is adopted to extract
relevant knowledge. The results generated by the generation model are optimized through the
discrimination module in the discriminative model of the generative adversarial network (GAN).

**Protection And Inheritance Of Ich**

**ICH features**

China's ICH is people's wisdom crystallization, with high historical and cultural value. The reason is that it
is the crystallization of culture and art produced in China's thousands of years of historical development,
and is the wealth handed down from history. Through them, people can understand the history, the
development level of social productivity in a specific historical period, the relationship between people in
the society at that time, the social structure and development mode, and the ideology, morality and
customs generated in the cultural environment at that time [8, 9]. It suggests that ICH accumulates the
essence of the times and national characteristics, is the inheritance of national history, reflects the
nation's survival state and world outlook, and reflects the nation's behavior mode and collective mentality.
It helps people make up for the lack of historical records, understand the society at that time, and help
people understand history more truly and comprehensively. The characteristics of ICH are summarized in
the following four points. 1. ICH is closely related to its surroundings and social environment. 2. ICH is not
unchangeable, with both inheritance and innovation. 3. Viability is a great feature of ICH. 4. ICH
inheritance has distinctive national and family style characteristics [10, 11]. Figure 1 displays some
typical ICHs:
Limitations of ICH image survival

ICH is an extremely rich and precious cultural wealth created by previous generations, and a crucial carrier of a nation's national spirit, national emotion, personality characteristics, cohesion and affinity. It contains oral works, national languages, folk performing arts, customs, etiquette, festivals, art music, musical instruments and traditional craft skills, all of which are closely related to the history of human culture. The science and technology progress and the globalization acceleration have led to the demise of multiple ICHs, some of which have disappeared. The current intervention of images plays a significant role in protecting ICH, especially the endangered heritage. However, it can also be seen that the survival of ICH images will also bring many negative effects, mainly because the images have certain limitations in recording ICH [12–14]. The first point is that the image recording time is limited, and the second point is that the image recording range is limited [15, 16]. Figure 2 presents the ICH protection framework.

The development of image restoration technology

Image repair refers to the use of technical means to completely display damaged image content. It is a popular study project in the current computer field. In recent years, due to the innovation as well as development of science and technology, image restoration technology has become a research project with great practical and scientific value. More development space and research projects can be explored through the research and development (R&D) of image restoration technology. Image restoration technology mainly has two types. First, the traditional computing method is adopted to supplement and predict the image content according to the image texture repair technology. Second, it is the repair technology based on computer Artificial Intelligence (AI) self-learning. The computer's AI neural network system is adopted for self-training of image repair technology [17, 18], and the computer measurement and neural system analysis mode are combined to achieve accurate prediction of image content. In the field of image restoration technology, the use of AI for the self-training of neural networks is a mainstream repair method. Although this repair method may not be able to achieve high-precision image texture repair, it can perform in-depth optimization on the image to gradually achieve an effective estimation of image information [19–21]. Figure 3 displays the development path of image restoration technology.

Figure 4 displays the current image repair effect. In the three groups of images shown in Fig. 4, the left image is damaged, and the right image is repaired.

Main Techniques Of Image Restoration Algorithm

Convolutional neural network (CNN) research

Being AI and machine learning's subset, DL algorithms greatly simplify the machine learning workflow. American scholars proposed the concept of DL in the second half of the 20th century.
DL initially aims to research the learning investment as well as knowledge mastery degree of learners. During learning, different learners will master knowledge through different learning methods. These methods are classified as DL as well as shallow learning. DL means learners' thinking, comprehending as well as putting forward their own problems during learning. Shallow-level learners will not pay attention to knowledge understanding. Instead, they will learn knowledge by remembering passively. Recently, DL has become a hot exploration direction in machine learning. It will enable the computer to study and conclude the internal laws of various data, and ultimately enable computers to obtain the analysis and learning capabilities similar to intelligent creatures. Currently, this technology has made multiple significant breakthroughs. In 2006, a new exploration direction named DL appeared in the machine learning, and started to be studied by the academic community and gradually applied by the industry [22, 23]. In 2012, Stanford University first established a training model called "deep neural network" by applying 16000 Central Processing Unit core parallel computing platform. It has made significant technical breakthroughs in the application fields such as voice as well as image recognition. In 2016, the artificial Go software established on the basis of DL defeated Li Shishi, who is the world class Go master. Later, various famous high-tech enterprises in the world started to put a lot of human and material in DL, set up DL research centres, and made massive technical R&D people enter the field of DL. Machine learning technology is to study the way that computers simulate and achieve animals' behavior of learning new knowledge or skills, rewrite the structure of data, and optimize the performance of the program. From the statistics perspective, machine learning is to forecast the distribution of data, obtain a model from the data and apply this model to new data prediction [24]. It is essential to ensure that there is the same distribution in the test as well as training data. The basic feature of machine learning is to make an attempt to imitate the mode of neuron information transmission as well as processing in human brain. Machine learning has a wide and prominent application in computer vision as well as natural language processing. It is obvious that DL has a close relationship with neural network in machine learning, and neural network is also the primary algorithm as well as means of DL. Alternatively, "DL" can be taken as the "improved neural network" algorithm. The modern deep CNN originates from the AlexNet network, which is the ancestor of deep CNN [25]. Compared with the previous CNN, the most significant feature of this network is its deeper level and larger parameter scale. Figure 5 displays its network structure:

Since AlexNet in 2012, CNN has a wide application in image classification and segmentation, target detection as well as other fields. Since there are increasingly higher performance requirements, AlexNet has been unable to satisfy everyone's needs, so researchers from all trades and professions have put forward CNN with better performance, like Visual Geometry Group (VGG), GoogLeNet, Residual Network (ResNet), and Densely Connected Network (DenseNet) properties. With the purpose of obtaining better performance, the network layer quantity will increase from layer 7 AlexNet to layer 16 VGG, from layer 16 VGG to layer 22 of GoogLeNet, to layer 152 ResNet and even thousands of layers of ResNet as well as DenseNet. The performance of network has been greatly improved, the efficiency problem remains to be solved. In recent years, researchers have paid close attention to the research of engineering technologies like model lightweight as well as compression. In these studies, there are several effective models:
SqueezeNet, MobileNet, and ShuffeNet. After summarizing its structure, it mainly uses bottleneck structure, grouping convolution structure as well as small size convolution kernel.

**Research on GAN**

GAN consists of a generative model (G) as well as a discriminative model (D). The generative model models the joint probability and describes the data generation process, and the discriminative model models the conditional probability to estimate the probability of samples from the training set. Here, the generative model aims to transform the input random noise into "music" by learning the true probability distribution of the data in the training set. The discriminative model mainly aims to correctly distinguish the "music" obtained from the generative model G from the music in the training set. The GAN is implemented by letting D as well as G play games. During the training process, D, together with G, is enhanced simultaneously through mutual competition. Since the discriminative model D exists, G can also learn features to approach the real data and there is no need to obtain lots of prior knowledge as well as prior distribution, so that the data obtained from G can finally realize the effect of mixing the spurious with the genuine.

The method of data processing here is to first divide an image into four equal parts, then use the relative brightness to encode each part of the image, and model each part separately. This exploration uses an array containing six lists to represent a image, as shown in (1):

$$(X_1, X_2, X_3, X_4, S, F)$$

1

In (1), $X_1, X_2, X_3, X_4$ represents four different parts in the image. $S$ represents brightness, and $F$ represents color.

To facilitate representation, this exploration assumes that there is only one image in the dataset, and defines a conditional probability distribution parameterized by $\theta_{i,t}$ as shown in Eq. (2):

$$\{p_{i,t} (V_i^t | V_i', S, F, q_{i,t})\}_{i \in [4], t \in [T]}$$

2

In (2), $[T]$ means the number of cycles, [4] means the part of the image that is divided, and $V_i'$ means other variables.

Here, a generative model $G^t$ consisting of four parts of neural networks is proposed to obtain a more accurate prediction structure. Figure 6 is a generative model diagram of one of the parts.

Part 1 of the proposed neural network model is composed of 3 Res-Desnet layers. It mainly aims to map the features at various time points to the space of the same dimension. After the mapping of the Res-Desnet layer, the outputs of various parts are connected into a sequence. Part 2 of this model is a two-
layer Bi-LSTM network. Part 3 and Part 4 of this model are Res-Desnet layers. The proposed discriminative model $D_i$ is composed of 4 convolutional layers, 2 Dense layers as well as 1 soft max layer. After each convolutional layer, Pre-RELU activation function is adopted to recombine features. Except for the first convolutional layer, each other convolutional layer contains a Batch normalization layer to improve the training performance. Figure 7 displays its network model. The convolution kernel quantity in the first, second, third, as well as fourth layers is set to 64, 128, 256 and 256, respectively. The convolution kernel size in the first, second, third, as well as fourth layers is same, and they are all 3×3. The step size in the first, second, third, as well as fourth layers is same and they are all 2.

**Dataset and model configuration**

Graphic Processing Unit (GPU) server is adopted for model training. The hardware configuration is Intel E52665X2, 32 GRECC DDR3, 250G solid state disk, as well as 4 NVIDIA RTX 2080TI 11G graphics cards. The software configuration is Ubuntu Linux 16.04, CUDA10.0, cuDNN7.6, and YOLOv3. The test is conducted on a laptop computer. The hardware configuration is Intel i79750H 4.5 GHz 6-core, with a memory of 32G DDR4 2666, as well as a GPU GeForce GTX 1650. The software configuration is Windows10, CUDA10.1, Cudnn7.6, OpenCV3.4.1.

**Analysis Of Experimental Results**

**Algorithm performance comparison**

Figure 8 displays the change curve of network error during the training of the unimproved GAN and the improved GAN.

The improved network training process reveals that when the number of iterations reaches 3500, the loss curves at 4550 and 6800 have small peaks, but the overall error of the network loss function displays a downward trend and tends to be flat. When the number of iterations reaches 8000, the loss function error of the network is basically stable below 0.2, and gradually tends to be stable. The original network training process shows that when the number of iterations reaches 7800, the loss function error of the network tends to be stable, basically below 0.25. The training error of the unimproved network is always higher than that of the optimized network, and the training effect of the optimized network model is superior to that of the original network model.

Table 1 shows the time consumption comparison of algorithms.
Table 1

<table>
<thead>
<tr>
<th>Scene</th>
<th>Resolution ratio</th>
<th>Unimproved GAN</th>
<th>Algorithm proposed here</th>
<th>Speedup</th>
</tr>
</thead>
<tbody>
<tr>
<td>Simple scenes</td>
<td>48×64×48</td>
<td>0.4825s</td>
<td>0.0391</td>
<td>12.2</td>
</tr>
<tr>
<td></td>
<td>96×128×96</td>
<td>5.0902s</td>
<td>0.361</td>
<td>14.7</td>
</tr>
<tr>
<td></td>
<td>192×256×192</td>
<td>58.2211s</td>
<td>3.361</td>
<td>18.1</td>
</tr>
<tr>
<td></td>
<td>384×512×384</td>
<td>65.5566s</td>
<td>4.451</td>
<td>20.2</td>
</tr>
<tr>
<td></td>
<td>768×1024×768</td>
<td>75.5566s</td>
<td>6.451</td>
<td>22.4</td>
</tr>
<tr>
<td>Complex scenes</td>
<td>48×64×48</td>
<td>0.4625s</td>
<td>0.0491</td>
<td>14.2</td>
</tr>
<tr>
<td></td>
<td>96×128×96</td>
<td>5.1902s</td>
<td>0.461</td>
<td>16.7</td>
</tr>
<tr>
<td></td>
<td>192×256×192</td>
<td>54.2211s</td>
<td>4.361</td>
<td>17.1</td>
</tr>
<tr>
<td></td>
<td>384×512×384</td>
<td>66.5566s</td>
<td>5.451</td>
<td>19.2</td>
</tr>
<tr>
<td></td>
<td>768×1024×768</td>
<td>78.5566s</td>
<td>7.451</td>
<td>25.4</td>
</tr>
</tbody>
</table>

In Table 1, in a simple scenario, when the resolution of the processed image is 48×64×48, it takes 0.4825s for the unimproved GAN to process the image, while the algorithm proposed here only needs 0.0391s to process the image, with a speedup of 12.2. In complex scenes, when the resolution of the processed image is 48×64×48, it takes 0.4625s for the unimproved GAN to process the image, while the algorithm proposed here only takes 0.0491s to process the image, with a speedup of 14.2. The results show that under different resolutions, the proposed algorithm has a speedup of more than 10 times compared with the unimproved GAN, and the advantages of the algorithm in the case of high resolution will be more obvious.

Figure 9 presents the prediction speed test results of the model built here.

Figure 9 suggests that in the system built here, the interaction accuracy of the second part of the segmented image is the highest, reaching 92%. The overall image recognition accuracy is in the range of 72%~92%, and the corresponding response time of the four parts of the segmented image is in the range of 5.1s-5.9s, which shows that the system built has a high real-time performance.

Comparison of algorithm display results

Figure 10 displays the comparison between the experimental results of the proposed algorithm and the experimental results of the unimproved GAN.

The comparison results in Fig. 10 show that under different resolutions, the results of the improved algorithm are very similar to those of the unimproved algorithm. Figure 10(b) suggests that the improved method can also well handle complex scenes with more details.
Conclusion

This exploration proposes a GAN to solve the problem of image restoration. A two-layer Bi-LSTM network with Attention mechanism is used as the generative model, and a CNN with four convolutional layers and two Dense connection layers is used as the discriminative model to generate image sequences. Then, the prediction accuracy is adopted to evaluate the model proposed. The experimental results suggests that the accuracy of the proposed model is superior to that of the previous model. However, the algorithm proposed has multiple shortcomings. The adjustment and control of GAN parameters proposed are poor. After each parameter adjustment, it is essential to train again. It is a waste of time. Thereby, the method that can ensure the image quality as well as stable performance will be the next research direction.

Declarations

Ethical Approval

No animal studies are presented in this manuscript.

No human studies are presented in this manuscript.

No potentially identifiable human images or data is presented in this study.

Conflict Of Interest(COI) statement

The authors has no conflict of interest.

Funding

No funding was used in this study.

DATA AVAILABILITY STATEMENT

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

Authors' contributions

The author contributed to manuscript revision, read, and approved the submitted version.

References


**Figures**
Figure 1

ICH

![Diagram](image)

Figure 2
ICH protection framework

Figure 3

Development path of image restoration technology

Figure 4
Image repair effect

Figure 5

CNN

Figure 6

Generative model $G'$
Figure 7

Discriminative Model $D'$

![Discriminative Model $D'$](image)

Figure 8

Variation curve of network error (a. Scenario 1 simple scenario; b. Scenario 2 complex scenario)
Figure 9

Test results of model prediction speed (a. The first test results; b. The second prediction results)
Figure 10

Comparison of experimental results of improved GAN and unimproved GAN (a. simple scenario; b. complex scenario)