International Journal for Housing Science and Its Applications
Publish August 10, 2025. Volume 46, Issue 4 Pages 2424-2433

https://doi.org/10.70517/ijhsa464205

# Three-dimensional reading and dynamic inheritance: Exploring the application of AR technology in the design of local cultural heritage education books

**Nan Ding**[1,*]
[1] College of Art and Design, Pingdingshan University, Pingdingshan, Henan, 467000, China
Corresponding authors: (e-mail: dingnan369@126.com).

**Abstract** The revitalization of local cultural heritage provides impetus for regional cultural development. This paper utilizes a cyclic threshold selection algorithm and a convolutional neural network (CNN) classification model to enhance the accuracy of 3D modeling and video animation scanning recognition of local cultural heritage educational books using augmented reality (AR) technology. The algorithm calculates an adaptive threshold for images, segments binary images, and combines the cosine loss function of image feature vectors to optimize the model's ability to correctly classify and differentiate between frames. AR scanning and recognition of local cultural heritage educational books are implemented on both the Vuforia and Unity platforms to enhance the immersive reading experience. Research findings indicate that when the number of pooling layers and convolutional layers are 4 and 3, respectively, the model achieves the lowest loss function value and the highest scores across all seven evaluation metrics. The average values for 4pt-Homography RMSE, PSNR, and SSIM are 0.5524, 30.7333, and 0.9365, respectively, with registration classification performance superior to that of the comparison model. The average scores of the experimental group students were significantly higher than those of the control group at the 0.01 level, and their satisfaction with the AR-based auxiliary teaching method reached 100%.

**Index Terms** Cultural heritage education books; AR technology; 3D modeling; image registration; CNN

## I. Introduction

Cultural heritage, as a testament to human history and civilization, has always been a focal point of concern for all sectors of society regarding its protection and preservation [1], [2]. With the rapid development of information technology, the theory of cultural heritage digitization has emerged, providing new perspectives and methods for cultural heritage protection [3]. As an important component of digital technology, AR (augmented reality) technology has taken this theory to new heights through its innovative applications in cultural heritage protection [4]-[6].

AR technology first emerged in the 1960s and has seen rapid development in recent years. By overlaying virtual objects onto the real world, AR technology enables users to experience virtual information within a specific spatial range, thereby achieving a sensory experience that transcends reality [7]-[10]. AR technology first entered the public eye through its use in broadcasting football matches, after which it experienced explosive growth, appearing in various industries. Many countries have also achieved significant results in combining cultural heritage with AR technology [11]-[13]. Chen, S., et al. explored the application of AR technology in cultural heritage tourism, quantitatively analyzed the impact of AR technology on visitors' cultural heritage responsibility behaviors, and further examined its influence on perceived usefulness, enjoyment, and behavioral intentions [14]. Ibiş, A., and Çakici developed an interactive AR application aimed at strengthening cultural heritage protection efforts, assessed its usability, and demonstrated its significant effectiveness in providing users with detailed historical information and enhancing user satisfaction [15]. Boboc, R., et al. provided a comprehensive overview of AR technology applications in the cultural heritage field over the past decade, identifying eight key themes—3D reconstruction, digital heritage, virtual museums, user experience, education, tourism, intangible cultural heritage, and gamification—as current research hotspots [16]. Xu, N, and others developed a mobile VR interactive exploration game called “Chinese Cultural Heritage Sites” based on AR technology and analyzed its impact on cultural heritage transmission and experience, finding that it has positive implications for advancing cultural heritage transmission and innovation [17].

AR technology can capture three-dimensional information about cultural heritage with unprecedented precision and detail, enabling comprehensive digital documentation of its form, structure, color, and texture. Andrade, J., and Dias, P. explored the impact of AR technology on the Regaleira cultural site, analyzing its advantages in enhancing visitor experience and increasing visibility and reputation [18]. Han, P, et al. proposed an immersive virtual reality

system based on AR technology to enable visitors to view fragile Dunhuang murals. This system utilizes digital data and interactive technology to deepen visitors' understanding of cultural heritage, making the murals clearly visible in a virtual scene, allowing visitors to freely explore the cultural treasures within the Mogao Grottoes [19]. González Vargas, J., et al. studied the impact of AR technology on student learning motivation in cultural heritage education. The study found that AR technology can enhance learning outcomes, stimulate learning motivation, and promote collaboration, thereby positively influencing student learning motivation in topics related to cultural heritage [20]. Garau, C explored the use of AR applications to enhance the cultural heritage experience on Sardinia. This initiative represents a unique attempt by the region to promote cultural heritage protection and address the issue of fragmented tourism products [21].

In terms of interactive learning, one of the advanced applications of AR technology in cultural heritage preservation is the design of interactive learning modules. Srdanović, P., et al. proposed integrating game elements, virtual reality (VR), and AR into cultural heritage transmission and preservation, developing a mobile application to enhance engagement and educational outcomes. The study found that this approach significantly outperformed traditional cultural heritage education in book form [22]. Bolognesi, C., et al. developed and implemented an AR application on a holographic table, aiming to achieve an immersive cultural heritage exploration experience that balances detail and performance [23]. Shih, N., et al. utilized AR technology in their research project to protect and promote traditional cultural elements and urban landscapes in the Lu Cang region of Taiwan by creating interactive digital models and providing remote on-site learning experiences [24]. Bekele, M., and Champion, E. compared immersive reality technologies and interaction methods to enhance cultural learning outcomes in virtual heritage applications. They proposed combining AR with multimodal interaction methods to establish contextual relationships, thereby increasing engagement between cultural contexts and virtual environments [25]. Gherardini, F., et al. integrated 3D virtual models with AR technology to overcome spatial and temporal constraints, applying this method to an exhibition of Roman artifacts in Modena, Italy, which received positive feedback from experts and visitors [26]. Pervolarakis, Z, and others utilized large-scale digitalization technology to create unique virtual tour experiences for historical and cultural heritage sites, specifically through VR and AR technologies, to enhance offline and on-site cultural heritage education and recreational experiences [27]. Comes, R, and others found that traditional local cultural heritage education, which relies on books for promotion and education, lacks interactivity. However, the interactivity and accessibility of AR technology offer unlimited potential for enhancing user engagement and understanding [28].

A summary of the above literature reveals that the ability to develop and create AR technology in conjunction with cultural heritage has significantly improved, while also stimulating people's enthusiasm for traditional cultural heritage. However, China is still in the exploratory stage of using AR technology to develop cultural heritage. Currently, some AR cultural and creative products merely apply AR technology to display cultural merchandise, so further research is needed.

This paper employs a cyclic threshold selection algorithm to achieve adaptive threshold calculation for video images, thereby enhancing the accuracy of target image and background image segmentation and precisely tracking visual image differences between frames. After segmenting the target image from the background, a registration model is selected to determine its differences from the actual image. By calculating the cosine loss between feature vectors, the computational burden on the model is reduced. To enhance image discrimination capabilities while improving classification accuracy, additive margin values (margin information) and additive margins are introduced to constrain image discrimination, and a serial-parallel mechanism is employed to accelerate classification computation efficiency. Vuforia and Unity platforms are utilized for the development of 3D models and video animations of local cultural heritage educational books, thereby enhancing AR scanning recognition accuracy and subsequent interactive immersion.

## II. Scanning and modeling of local cultural heritage education books supported by AR technology

### *II. A. Visual image detection method based on interframe differences*

Augmented reality technology is inextricably linked to visual image detection. To identify markers in the real world, devices must perform visual image detection on the acquired video images to extract the markers. There are numerous visual image detection methods, each with its own advantages and disadvantages, as well as specific applications and directions. Current detection methods include background subtraction, frame-to-frame difference, and optical flow. The inter-frame difference method is particularly suitable for AR systems on handheld devices, and we will focus on this method here.

The inter-frame difference method is a technique that calculates the difference between adjacent frames in a video image sequence to obtain the outline of a moving object. It is well-suited for scenarios involving multiple moving objects and camera movement. When moving objects appear in the background scene, the frames captured by the camera exhibit noticeable differences. By subtracting the gray-scale values of each pixel between two frames,

the gray-scale difference between the two frames is obtained. The difference image is then binarized: if the gray-scale difference exceeds a predefined threshold, the pixel is classified as a target pixel; otherwise, it is classified as a background pixel. By determining whether there are moving objects in the captured images through the binary image, the sequential frames of the image sequence are compared frame by frame, which is equivalent to applying a high-pass filter to the image sequence in the time domain.

The key to determining moving targets and background in the binary image lies in the setting of the threshold. Adaptive thresholds can significantly improve the accuracy of the determination and reduce the influence of background environmental lighting and shadows. An adaptive threshold selection algorithm can be used to achieve adaptive threshold determination. The specific algorithm steps are as follows:

First, take the maximum gray value $D_1$ and $D_k$ from the binarized image, and initialize the threshold as:

$$T_0 = \frac{G_1 + G_k}{2} \tag{1}$$

Then, based on the threshold $T_k$ obtained from the previous calculation, divide the image into two parts and calculate the average gray value of each part.

$$G_A = \frac{\sum_{G(i,j)<T_k} G(i,j) \times D(i,j)}{\sum_{G(i,j)<T_k} D(i,j)} \tag{2}$$

$$G_B = \frac{\sum_{G(i,j)>T_k} G(i,j) \times D(i,j)}{\sum_{G(i,j)>T_k} D(i,j)} \tag{3}$$

where $G_A$ is the average grayscale value of the moving target, and $G_B$ is the average grayscale value of the background. $G(i,j)$ is the grayscale value of the pixel point, and $D(i,j)$ is the weight coefficient of the corresponding pixel point, which is generally 1. The algorithm can also be optimized by setting the weight values of each point to improve accuracy.

The new threshold is then calculated using the average grayscale values of the moving target and the background obtained from the above calculations.

$$T_{k+1} = \frac{G_A + G_B}{2} \tag{4}$$

If $T_k = T_{k+1}$, the loop ends; otherwise, the iteration continues. After the calculation ends, the obtained threshold is used to binarize the image. The binarized target image still contains a large amount of noise, so subsequent processing such as morphological filtering and connectivity analysis is required to increase the accuracy of the target segmentation.

### *II. B.Design of AR registration system based on loss function*

#### II. B. 1) Loss Design

Visual image alignment in local cultural heritage education books often uses distance as the basis for loss design, with the aim of shortening the distance between aligned sample feature vectors and expanding the space of unaligned samples. The images to be aligned are matched with the output of a twin neural network in the form of multidimensional feature vectors, after which formulas (5) and (6) are used to output the corresponding Euclidean distance or L2 distance:

$$E_k = \frac{1}{2}\sum_{j=1}^{n}(C_{x_j} - C_{y_j})^2 \tag{5}$$

$$L_c = \frac{1}{2}\sum_{i=1}^{n}\left\|C_{x_i} - C_{y_i}\right\|_2^2 \tag{6}$$

Here, $n$ denotes the number of output features, and $C_x, C_y$ correspond to the feature sets output by the twin networks $G(I_x)$ and $G(I_y)$ respectively. When comparing the registration algorithm with the Center-Loss in the field of face recognition, it is found that the objects being measured are completely different. The loss calculated by the registration algorithm is the distance between the two features extracted by the twin networks, while the loss calculated by Center-Loss is the distance between the label and the CNN output features. Similarly, combining the

cosine loss from the face recognition field, such as LMCL, L-Softmax, and A-Softmax, the cosine loss is based on the Softmax loss, calculating the cosine value between the weights of the last fully connected layer and the CNN output features. Inspired by traditional distance loss and face recognition loss, this paper calculates the cosine value between the two output feature vectors of the Pseudo-Siamese network and uses this as the basis for loss design.

Cosine loss retains the advantage of amplifying inter-class differences while reducing sensitivity to individual signals, focusing more on directional differences. The magnitude of inter-class differences is assessed based on the cosine similarity between two feature vectors. Cosine similarity is defined as follows:

$$\cos\theta_l = \frac{C_x * C_y}{\| C_x \| * \| C_y \|} = \frac{\sum_{i=1}^{n} C_{x_i} * C_{y_i}}{\sqrt{\sum_{i=1}^{n} C_{x_i}^2} * \sqrt{\sum_{i=1}^{n} C_{y_i}^2}} \tag{7}$$

where $\theta_l$ is the angle of the feature vector, $\cos\theta_l$ is the range of feature similarity [-1.0,1.0], if the input image matches, the angle $\theta_l$ is smaller $\cos\theta_l$ is closer to 1.0, if the input image does not match, the angle is larger $\cos\theta_l$ is closer to -1.0, $n$ Represents the feature output dimension. In the actual calculation, it is necessary to output a cost function that can distinguish distances, such as L2-Net design error loss and Triplet loss. In this paper, the loss function is designed according to the cosine similarity as follows:

$$L_l = \frac{1}{2} l(s * cos\,\theta_l - 1)^2 + \frac{1}{2}(1-l)(s * cos\,\theta_l + 1)^2 \tag{8}$$

During the loss calculation process, we need to ensure that only the angle between the feature vectors is obtained, independent of the feature weights. Therefore, the L2 algorithm is used to normalize the individual feature weights. When calculating the recognition score for an image pair, only the cosine value $cos\theta_l$ of the two feature vectors is required, indicating that the norms of $C_x, C_y$ do not contribute to the score. Finally, the repair parameter $s$ is given. $l$ is a binary label used to determine whether the input image pair $I_x, I_y$ is a positive sample $(l = 1.0)$ or a negative sample $(l = 0.0)$. It can be seen that the cosine loss is independent of feature strength. Therefore, using the cosine loss as the loss function for the registration model can transform the loss from Euclidean space to angular space, thereby eliminating the influence of signal strength and reducing the burden on the network.

#### II. B. 2) Margin Design

Since cosine loss only emphasizes correct classification, the features learned by the twin network in the test do not have sufficient discriminative power. To solve this problem, this paper introduces additive margin values (margin information) into the cosine similarity. Figure 1 shows the cosine angle structure. Since the cosine angle is distributed between [0.0, $\pi$], we add an additional angle margin penalty $m$ between the feature vectors $C_x, C_y$ of the original image and the image to be aligned, thereby enhancing intra-class compactness and inter-class differences. If the inner product between the feature vectors learned by the twin network is $cos\theta_j$, the cosine angle classification threshold is set to $v$. If the input images match, then $cos\theta_l \in (v, 1.0]$; if they do not match, then $cos\theta_l \in [-1.0, v]$, ensuring that features with different attributes are correctly classified. To develop an edge classifier, we further require that positive samples $\cos\theta_l \in (v+m, 1.0]$ and negative samples $cos\theta_j \in [-1.0, v-m]$. Here, $m$ controls the cosine margin size, with a value range of $m \in [0.0, 1, 0]$. Since the sets $(v+m, 1.0] \in (v, 1.0], [-1.0, v-m] \in [-1.0, v]$, adding the margin makes the loss more stringent in constraining the classification.

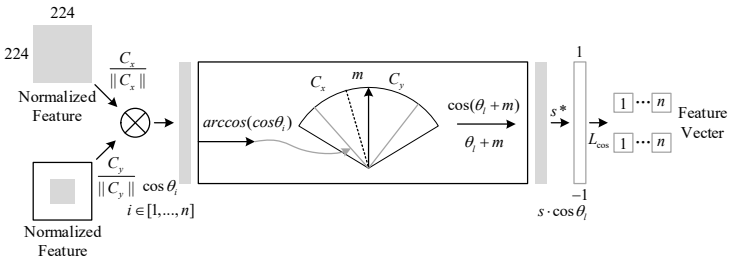


Figure 1: Illustration of the cosine Angle

The additive angular margin penalty proposed in this paper is equivalent to the normalized distance margin penalty. The additional angular margin cosine loss retains the advantage of expanding inter-class differences but reduces sensitivity to different signal strengths, placing greater emphasis on the differences in direction vectors. More precisely, it encourages similar feature vectors to have as small an angle as possible, while different feature vectors have as large an angle as possible. Formally, combining the LMCL loss, the additional angular margin cosine loss function (AM-Cos) is defined as follows:

$$L_{cos} = \frac{1}{2N}\sum_{j=1}^{N}\left[ l(cos(\theta_{l_j} + m) - 1)^2 + (1-l)(cos(\theta_{l_j} - m) + 1)^2 \right] \tag{9}$$

In the equation, $N$ is the training sample number, and $\theta_{l_i}$ is the angle between the $j$ th feature vector output by the twin network. In the feature angle loss function, the compensation value $s$ has almost no contribution to feature determination and has no effect on the final loss calculation result, so $s = 1.0$ is labeled. The paper adopts an additive margin addition method to weaken the marginal penalty, using the inverse cosine function to add the margin value, thereby avoiding the increased training difficulty caused by the multiplicative marginal penalty and the complex double-angle formula.

#### II. B. 3) Attention Parallel Mechanism Sorting

In the AR registration system, to make full use of contextual information, we combine anti-rotation mechanisms with parallel mechanisms. Specifically, we transform feature information through pooling layers and accumulate different weights for pixels or channels through various forms of pooling layers to model the importance of features. Finally, we output the final prediction map through different residual blocks. Through experiments, we found that arranging the two mechanisms in series yields better results than arranging them in parallel. In the specific model, we adjust the positions and quantities of the two mechanisms throughout the model to improve the computational efficiency of the GPU-accelerated CNN library.

### *II. C.AR scanning and recognition function*

In the development of immersive interactive features, AR scanning and recognition functionality is a key focus of the project and serves as the foundation for subsequent feature development. In the development of AR recognition functionality, this research project primarily focuses on AR recognition of 3D models from local cultural heritage education books and AR recognition of video animations. The AR recognition functionality development utilizes Vuforia integrated with Unity as the development tools. First, AR development permissions must be obtained on the Vuforia side. Then, recognition development for the 3D models of local cultural heritage education books and video animations must be conducted within Unity 3D. The following sections detail the specific operations on the Vuforia and Unity sides.

#### II. C. 1) Vuforia End

First, log in to the Vuforia official website and register an account to apply for a License Key. In the Develop menu, select License Manager and add a License Key. The License Key grants project development engineering permissions, enabling developers to access resources in the Vuforia SDK.

Next, create a database for recognition objects. Use the Target Manager to create the database and add the prepared recognition images. This allows the recognition images for Image Targets to be saved via the Vuforia server. Finally, click to download the database and generate a Unitypackage recognition image resource package as a backup. When developing in Unity, this package must be imported into the project.

#### II. C. 2) Unity End

First, import the files. In the Unity project panel, delete the Main Camera from the original scene and add an AR Camera. Import the Unitypackage recognition image resource package generated on the Vuforia end, as well as the 3D models and other material resources required for scanning and recognition.

Next, configure the parameters and settings. When setting up the AR Camera, enter or paste the license key obtained and saved in advance from the Vuforia side into the Vuforia Behavior of the AR Camera to enable scanning permissions for the recognition image. Check the upload status of the recognition image database in the Database Load Behavior, select Load Database to complete activation. Next, set up the Image Target. Add the Image Target to the editing panel and adjust its position to appear in the center of the AR camera. Add the recognition image data to the Image Target Behavior.

Finally, adjust the imported 3D models, scene assets, and video animation assets of local cultural heritage educational books. For example, when adjusting the 3D model, add the 3D model to the editing scene, drag it into the Image Target, and set it as a child object of the Image Target. Use the move and rotate tools to adjust the model's position and angle so that it is centered on the screen. After adjusting the position, set the model to invisible

and use a script to control its visibility. Place the model in the “Prefer” folder under the “Resource” menu and set it as a prefab. Add a script-attached master controller component to enable asynchronous loading of the model.

## III. Model performance verification and application analysis based on AR technology

### *III. A. Analysis of the model training process and optimal parameter settings*

#### III. A. 1) Analysis of experimental results before model enhancement

For the constructed CNN models, we conducted multiple experiments, statistically analyzed the experimental results of each prediction during model testing, and calculated the specific values of each model's performance metrics. By comparing each metric, we evaluated the strengths and weaknesses of each group of CNN models and selected the optimal classification and registration model by comparing the differences among the experimental models. Table 1 summarizes the performance metrics of the models across multiple experiments. Observing the experimental results, we can conclude that when the number of pooling layers is 3 and the number of convolutional layers is 2 (i.e., CNN model 5), it achieves the highest sensitivity (0.9033). CNN model 2 achieves the best result in the mCC metric (0.9176). CNN model 3 has the highest Acc of 0.9085. CNN model 4 achieves the highest FMI of 0.9188. Through a comprehensive analysis, we can conclude that different models exhibit differences in their metric results. Therefore, further comparison with data augmentation and loss functions is necessary to determine the final model parameter settings for AR scan recognition.

Table 1: Statistics of model index results

| Serial number | Sen | Spc | Prc | Acc | F1 | mCC | FMI |
|---|---|---|---|---|---|---|---|
| 1(1,2) | 0.8973 | 0.9235 | 0.8775 | 0.8377 | 0.8181 | 0.8861 | 0.8917 |
| 2(1,3) | 0.8549 | 0.8948 | 0.8148 | 0.8183 | 0.8567 | 0.9176 | 0.8185 |
| 3(2,2) | 0.6744 | 0.7754 | 0.8743 | 0.9085 | 0.7843 | 0.7427 | 0.8012 |
| 4(2,3) | 0.6952 | 0.7293 | 0.8573 | 0.7908 | 0.8415 | 0.8404 | 0.9188 |
| 5(3,2) | 0.9033 | 0.8317 | 0.7879 | 0.7758 | 0.8078 | 0.8302 | 0.8284 |
| 6(3,3) | 0.7725 | 0.8688 | 0.8836 | 0.7898 | 0.8155 | 0.8109 | 0.8246 |
| 7(4,2) | 0.8776 | 0.7701 | 0.7427 | 0.7347 | 0.9063 | 0.8624 | 0.6693 |
| 8(4,3) | 0.7368 | 0.8019 | 0.9131 | 0.7675 | 0.7029 | 0.8272 | 0.8481 |

#### III. A. 2) Analysis of experimental results after model enhancement

The differences in the data for each indicator in Table 1 can reflect the performance of the model from different perspectives. We can see that no single model excels in all aspects, and there are subtle differences between the models. However, we need to identify a relatively stable and excellent model to best accomplish our classification task. After data augmentation processing, the size of our dataset has significantly increased, expanding from the original 4,500 samples to 45,000, with the number of data samples increasing tenfold. This allows for the inclusion of more video image scenes from local cultural heritage education books, which is of great significance for enhancing the model's generalization ability.

In further experiments, we recorded the convergence of each model's loss function to observe the training process. Figure 2 shows the convergence process of multiple CNN models. In the 8 CNN model experiments, the loss functions converged rapidly to very low levels, indicating that our experimental models quickly reached stability, and data augmentation indeed played a crucial role. Among them, the training loss function values of the eighth model, CNN8, were all below 0.025, and the function fluctuations were the most stable.

Additionally, to compare with the previous group of experimental models without data augmentation, we also recorded the seven metrics from Section 3.1.1 to analyze the data results. Table 2 shows the metric data from the experiments with data augmentation. When the number of pooling layers is 4 and the number of convolutional layers is 3, the seven metric values reach 0.9521, 0.9535, 0.9369, 0.9379, 0.9468, 0.9436, and 0.9295, respectively, all at the highest levels, indicating the best model registration classification performance. Therefore, this paper sets the number of pooling layers in the CNN model to 4 and the number of convolutional layers to 3 to improve the AR scanning recognition and image registration classification accuracy of 3D models of local cultural heritage education books, etc.

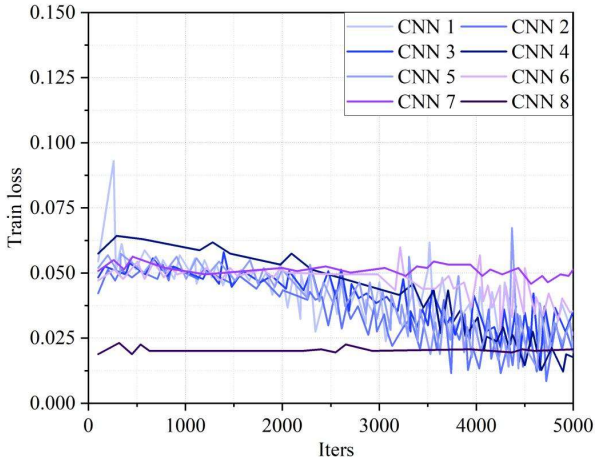


Figure 2: The convergence process of CNN model training

Table 2: The model combines data to enhance the analysis of indicators

| Serial number | Sen | Spc | Prc | Acc | F1 | mCC | FMI |
|---|---|---|---|---|---|---|---|
| 1(1,2) | 0.9263 | 0.8404 | 0.8982 | 0.8671 | 0.8481 | 0.9022 | 0.9019 |
| 2(1,3) | 0.8838 | 0.9149 | 0.8355 | 0.8475 | 0.8872 | 0.8534 | 0.8288 |
| 3(2,2) | 0.8036 | 0.8055 | 0.8975 | 0.8168 | 0.8148 | 0.8281 | 0.8115 |
| 4(2,3) | 0.8242 | 0.8594 | 0.8738 | 0.8203 | 0.8745 | 0.8665 | 0.8584 |
| 5(3,2) | 0.8657 | 0.8818 | 0.8086 | 0.8052 | 0.8383 | 0.8563 | 0.8497 |
| 6(3,3) | 0.8014 | 0.8989 | 0.9043 | 0.8196 | 0.8426 | 0.8437 | 0.8348 |
| 7(4,2) | 0.9065 | 0.8004 | 0.7634 | 0.8644 | 0.8334 | 0.8855 | 0.7796 |
| 8(4,3) | 0.9521 | 0.9535 | 0.9369 | 0.9379 | 0.9468 | 0.9436 | 0.9295 |

### *III. B. Comparison of registration results between different models*

#### III. B. 1) Comparison of 4pt-Homography RMSE Errors

This section compares the designed registration classification model with other methods to determine its registration classification advantages, using objective data metrics as the comparison standard. The comparison methods selected are SIFT+RANSAC, dhN, UDHN, CA-UDHN, and UDIS. The model is trained using the Warped-COCO dataset, and the 4pt-Homography RMSE objective evaluation metric is used for comparison. Subsequently, the model is further trained using the UDIS-D dataset, and the PSNR and SSIM objective evaluation metrics are used for comparison. To provide a more intuitive comparison of the model's registration performance, this paper also calculated objective metrics for the results obtained by transforming images using the identity matrix I3×3 without using a homography matrix. Table 3 summarizes the 4pt-Homography RMSE error comparison results of the model on the Warped-COCO dataset. On the synthetic dataset Warped-COCO, the 4pt-Homography RMSE of this paper's model achieved the smallest error values across the three data ranges of 0-35%, 35-65%, and 65-100%, with values of 0.4561, 0.5729, 0.6283, respectively, with an average error of only 0.5524, outperforming the comparison models in registration error.

Table 3: 4pt-Homography RMSE error comparison

| | Top 0-35% | Top 35-65% | Top 65-100% | Average |
|---|---|---|---|---|
| $I_{3\times3}$ | 14.9033 | 17.2451 | 18.1647 | 16.7710 |
| SIFT+RANSAC | 0.8467 | 0.9518 | 0.9934 | 0.9307 |
| DhN | 2.7945 | 3.1263 | 2.1158 | 2.6788 |
| UDHN | 1.4729 | 2.5328 | 3.6420 | 2.5492 |
| CA-UDHN | 3.4710 | 4.1745 | 3.5681 | 3.7379 |
| UDIS | 4.3528 | 3.4680 | 2.3629 | 3.3946 |
| Article method | 0.4561 | 0.5729 | 0.6283 | 0.5524 |

#### III. B. 2) PSNR Comparison

Table 4 summarizes the peak signal-to-noise ratio (PSNR) comparison results of the model on the UDIS-D dataset. The PSNR of the model in this paper reached 30.4663, 31.6189, and 30.1187 at three levels, with an average value

of 30.7333, which was the highest level in the model comparison. The images classified by the model in this paper had little difference from the original images and had better registration quality.

Table 4: PSNR comparison of the model on the UDIS-D dataset

| | Top 0-35% | Top 35-65% | Top 65-100% | Average |
|---|---|---|---|---|
| $I_{3\times3}$ | 15.2679 | 14.1579 | 12.3765 | 13.8769 |
| SIFT+RANSAC | 23.1584 | 21.4317 | 20.2589 | 21.6727 |
| DhN | 15.3105 | 12.6728 | 11.8434 | 13.3859 |
| UDHN | 25.1203 | 21.4890 | 20.6394 | 22.4242 |
| CA-UDHN | 17.2475 | 15.2831 | 16.3167 | 16.2448 |
| UDIS | 20.3682 | 17.3274 | 15.6904 | 17.9134 |
| Article method | 30.4663 | 31.6189 | 30.1187 | 30.7333 |

**III. B. 3) SSIM Comparison**

Table 5 summarizes the structural similarity index (SSIM) comparison results of the model on the UDIS-D dataset. In terms of SSIM, the indicator value of the model in this paper is very close to 1.0000, with indicator values of 0.9546, 0.9384, and 0.9201 at three levels, averaging 0.9365, all of which are higher than those of the comparison model. The two images aligned by the proposed model are nearly identical, demonstrating excellent image quality.

Table 5: SSIM comparison of the model on the UDIS-D dataset

| | Top 0-35% | Top 35-65% | Top 65-100% | Average |
|---|---|---|---|---|
| $I_{3\times3}$ | 0.5728 | 0.3961 | 0.1203 | 0.3600 |
| SIFT+RANSAC | 0.6384 | 0.5069 | 0.4671 | 0.5426 |
| DhN | 0.5940 | 0.2630 | 0.2482 | 0.3786 |
| UDHN | 0.6521 | 0.4573 | 0.5268 | 0.5450 |
| CA-UDHN | 0.6481 | 0.7410 | 0.6923 | 0.6900 |
| UDIS | 0.4220 | 0.3415 | 0.2097 | 0.3267 |
| Article method | 0.9546 | 0.9384 | 0.9201 | 0.9365 |

### *III. C. Stereoscopic reading comparison experiment of educational books based on AR technology*

**III. C. 1) Comparison of scores before and after the experiment**

An AR-based local cultural heritage education book was applied to the general education courses for sophomores at a university, and a control experiment was set up to study the effectiveness of these books in cultivating students' reading abilities and awareness of cultural heritage. The experiment selected 200 sophomores, divided into two groups of 100 each: an experimental group and a control group. The experimental group used AR-based local cultural heritage education books as learning materials, while the control group used conventional books. The experiment lasted one semester (14 weeks), with pre- and post-tests of reading ability and heritage awareness conducted at the beginning and end of the experiment. Table 6 shows the comparison of test scores before and after the experiment. The P-value for the total scores of the two groups before the experiment was 0.109, greater than 0.05, indicating no significant difference, which meets the requirements for a control experiment. The P-value after the experiment was 0.005, less than 0.01, indicating that after the control experiment, the total scores of the two groups showed a significant difference at the 0.01 level.

Table 6: Comparison of test scores before and after the experiment

| | Group | N | M | SD | P |
|---|---|---|---|---|---|
| Before the experiment | Experimental group | 100 | 69.093 | 2.175 | 0.109 |
| | Control group | 100 | 68.838 | 2.175 | |
| After the experiment | Experimental group | 100 | 90.276 | 4.896 | 0.005 |
| | Control group | 100 | 72.915 | 2.054 | |

Table 7 shows the changes in reading ability and heritage awareness scores between the two groups of students before and after the experiment. In terms of reading ability, the average score of the experimental group increased from 68.219 to 90.283, and the number of students scoring 85 or above increased from 10 to 45. In terms of heritage awareness, the average score of the experimental group increased from 69.967 to 90.269, and the number of students scoring 85 or above increased from 11 to 50. After the experiment, the P-values for the two ability indicators

between the experimental group and the control group were 0.002 and 0.003, respectively, both less than 0.01, indicating that the experimental group was significantly superior to the control group at the 0.01 level. The application of AR technology-based local cultural heritage education books in teaching can enhance students' three-dimensional reading and living heritage awareness, as well as their learning abilities.

Table 7: Changes in students' reading ability and awareness of inheritance

| | Capability indicators | Group | N | M | Number of people | | P |
|---|---|---|---|---|---|---|---|
| | | | | | <85 | ≧85 | |
| Before the experiment | Reading ability | Experimental group | 100 | 68.219 | 90 | 10 | 0.108 |
| | | Control group | 100 | 68.897 | 89 | 11 | |
| | Inheritance consciousness | Experimental group | 100 | 69.967 | 89 | 11 | 0.110 |
| | | Control group | 100 | 68.779 | 90 | 10 | |
| After the experiment | Reading ability | Experimental group | 100 | 90.283 | 55 | 45 | 0.002 |
| | | Control group | 100 | 70.164 | 85 | 15 | |
| | Inheritance consciousness | Experimental group | 100 | 90.269 | 50 | 50 | 0.003 |
| | | Control group | 100 | 75.666 | 86 | 14 | |

**III. C. 2) Comparison of teaching satisfaction**

A questionnaire survey was conducted to assess the satisfaction levels of 100 students in the experimental group regarding this teaching method. A total of 100 questionnaires were distributed, and 100 were returned, resulting in a 100% response rate. Table 8 presents the statistical results of student satisfaction. In the five evaluation criteria—enhancing classroom engagement, excellent interactive effects, improving learning efficiency, strengthening three-dimensional reading skills, and promoting awareness of living heritage—student satisfaction rates reached 91.28%, 91.35%, 91.67%, 94.50%, and 94.26%, respectively. The combined rates of “very satisfied” and “satisfied” reached 100% in all categories. This indicates that utilizing AR technology to scan and identify local cultural heritage educational books and applying them to classroom teaching can enhance students' abilities while increasing classroom engagement and interactivity, thereby fostering greater student involvement in learning.

Table 8: The statistical results of students' satisfaction

| Evaluation index | Proportion (%) | | | | |
|---|---|---|---|---|---|
| | Very satisfied | Satisfied | General | Dissatisfied | Very dissatisfied |
| Increase the interest of the class | 91.28 | 8.72 | 0.00 | 0.00 | 0.00 |
| Excellent of interactive effect | 91.35 | 8.65 | 0.00 | 0.00 | 0.00 |
| Improve learning efficiency | 91.67 | 8.33 | 0.00 | 0.00 | 0.00 |
| Enhance the ability of stereoscopic reading | 94.50 | 5.50 | 0.00 | 0.00 | 0.00 |
| Promote the awareness of living inheritance | 94.26 | 5.74 | 0.00 | 0.00 | 0.00 |

## IV. Conclusion

This paper utilizes AR technology to scan and recognize local cultural heritage education books and explores their practical application value. The 4pt-Homography RMSE error of the model is 0.4561, 0.5729, and 0.6283 at three levels, respectively, with relatively small errors. The PSNR values reached 30.4663, 31.6189, and 30.1187, while the SSIM values reached 0.9546, 0.9384, and 0.9201, indicating superior image registration performance compared to the reference model. After the control experiment, there were significant differences in academic performance between the two groups at the 0.01 level. Additionally, the experimental group's reading ability and heritage awareness both improved to an average of over 90 points, with the number of outstanding students (≥85 points) increasing to 45 and 50, respectively. The experimental class students' very high satisfaction with AR-assisted teaching was 91.28%, 91.35%, 91.67%, 94.50%, and 94.26%, respectively. In the future, lightweight modules can be introduced to accelerate AR scanning and recognition efficiency and enhance interaction real-time performance.

## References


[1] Dascălu, I. (2018). Cultural Heritage and the Inheritance of Civilization. Insights from Michael Oakeshott's Reflections on Education. Hermeneia, (21), 133-144.

[2] Harvey, D. C. (2016). The history of heritage. In The Routledge research companion to heritage and identity (pp. 19-36). Routledge.

[3] Gervasi, O., Perri, D., Simonetti, M., & Tasso, S. (2022, July). Strategies for the digitalization of cultural heritage. In International Conference on Computational Science and Its Applications (pp. 486-502). Cham: Springer International Publishing.

[4] Zhao, Z. (2017, October). Digital protection method of intangible cultural heritage based on augmented reality technology. In 2017 International Conference on Robots & Intelligent System (ICRIS) (pp. 135-138). IEEE.
[5] Cantoni, V., Mosconi, M., & Setti, A. (2019, July). Technological innovation and its enhancement of cultural heritage. In 2019 IEEE International Symposium on Innovations in Intelligent SysTems and Applications (INISTA) (pp. 1-6). IEEE.
[6] Wei, W. A. N. G., & Xin, X. U. (2024). Transformation and Development of Intangible Cultural Heritage through Technology. Journal of Library & Information Science in Agriculture, 36(1).
[7] Carmigniani, J., Furht, B., Anisetti, M., Ceravolo, P., Damiani, E., & Ivkovic, M. (2011). Augmented reality technologies, systems and applications. Multimedia tools and applications, 51, 341-377.
[8] Chen, Y., Wang, Q., Chen, H., Song, X., Tang, H., & Tian, M. (2019, June). An overview of augmented reality technology. In Journal of Physics: Conference Series (Vol. 1237, No. 2, p. 022082). IOP Publishing.
[9] Palmarini, R., Erkoyuncu, J. A., & Roy, R. (2017). An innovative process to select Augmented Reality (AR) technology for maintenance. Procedia Cirp, 59, 23-28.
[10] Syed, T. A., Siddiqui, M. S., Abdullah, H. B., Jan, S., Namoun, A., Alzahrani, A., ... & Alkhodre, A. B. (2022). In-depth review of augmented reality: Tracking technologies, development tools, AR displays, collaborative AR, and security concerns. Sensors, 23(1), 146.
[11] Ivanova, A. V. (2018). VR & AR technologies: opportunities and application obstacles. Strategic decisions and risk management, (3), 88-107.
[12] Zhao, L., & Kim, J. (2024). The impact of traditional Chinese paper-cutting in digital protection for intangible cultural heritage under virtual reality technology. Heliyon, 10(18).
[13] Liu, Z. (2023). Regeneration of Liangzhu culture: multimedia exhibition, simulated restoration, innovative cultural products, nearby area integration, virtual reality and augmented reality. Humanities and Social Sciences Communications, 10(1), 1-18.
[14] Chen, S., Tian, Y., & Pei, S. (2024). Technological use from the perspective of cultural heritage environment: Augmented reality technology and formation mechanism of heritage-responsibility behaviors of tourists. Sustainability, 16(18), 8261.
[15] Ibiş, A., & Çakici Alp, N. (2024). Augmented reality and wearable technology for cultural heritage preservation. Sustainability, 16(10), 4007.
[16] Boboc, R. G., Băutu, E., Gîrbacia, F., Popovici, N., & Popovici, D. M. (2022). Augmented reality in cultural heritage: an overview of the last decade of applications. Applied Sciences, 12(19), 9859.
[17] Xu, N., Li, Y., Liang, J., Shuai, K., Li, Y., Yan, J., ... & Dong, Y. (2024). HeritageSite AR: Design and Evaluation of a Mobile Augmented Reality Exploration Game for a Chinese Heritage Site. ACM Journal on Computing and Cultural Heritage, 17(4), 1-29.
[18] Andrade, J. G., & Dias, P. (2020). A phygital approach to cultural heritage: Augmented reality at regaleira. Virtual Archaeology Review, 11(22), 15-25.
[19] Han, P. H., Chen, Y. S., Liu, I. S., Jang, Y. P., Tsai, L., Chang, A., & Hung, Y. P. (2019). A compelling virtual tour of the dunhuang cave with an immersive head-mounted display. IEEE Computer Graphics and Applications, 40(1), 40-55.
[20] González Vargas, J. C., Fabregat, R., Carrillo-Ramos, A., & Jové, T. (2020). Survey: Using augmented reality to improve learning motivation in cultural heritage studies. Applied Sciences, 10(3), 897.
[21] Garau, C. (2014). Smart paths for advanced management of cultural heritage. Regional Studies, Regional Science, 1(1), 286-293.
[22] Srdanović, P., Skala, T., & Maričević, M. (2025). InHeritage—A Gamified Mobile Application with AR and VR for Cultural Heritage Preservation in the Metaverse. Applied Sciences (2076-3417), 15(1).
[23] Bolognesi, C. M., Sorrenti, D., & Bassorizzi, D. (2023). Utilizing stereography to compare cultural heritage in the past and now: an interactive AR application. Applied Sciences, 13(15), 8773.
[24] Shih, N. J., Chen, H. X., Chen, T. Y., & Qiu, Y. T. (2020). Digital preservation and reconstruction of old cultural elements in augmented reality (AR). Sustainability, 12(21), 9262.
[25] Bekele, M. K., & Champion, E. (2019). A comparison of immersive realities and interaction methods: Cultural learning in virtual heritage. Frontiers in Robotics and AI, 6, 91.
[26] Gherardini, F., Mattia, S., & Leali, F. (2019). Enhancing heritage fruition through 3D virtual models and augmented reality: An application to Roman artefacts. Virtual Archaeology Review, 10(21), 67-79.
[27] Pervolarakis, Z., Zidianakis, E., Katzourakis, A., Evdaimon, T., Partarakis, N., Zabulis, X., & Stephanidis, C. (2023). Visiting heritage sites in AR and VR. Heritage, 6(3), 2489-2502.
[28] Comes, R., Neamtu, C., Buna, Z., & Mateescu-Suciu, L. (2019). Exploring Dacian cultural heritage with dARcit augmented reality application. Journal of Ancient History and Archaeology, 6(4).