Sensors and Materials, Vol. 38, No. 2 (2026) 965–983
MYU Tokyo

S & M 4358

Teaching Traditional Embroidery Using Digital Immersive Tool Based on Augmented Reality and Sensor-based Recognition

Xuehong Zhao,[1] Mingyu Zhao,[2*] and Hailing Wang[3]

[1]Harbin Finance University, Harbin 150030, China
[2]Heilongjiang Academy of Sciences, Harbin 150090, China
[3]Heilongjiang University of Science and Technology, Harbin 150022, China

(Received June 5, 2025; accepted February 24, 2026)

Keywords: AR, Kinect/Leap Motion sensor, sensor fusion, embroidery, intangible cultural heritage

In this study, we introduced an augmented reality and deep-learning-based system for the digital preservation and interactive learning of traditional embroidery skills. Recognizing the vulnerability of traditions in preserving intricate intangible cultural heritage, this system integrates multisource data from Kinect, Leap Motion, inertial measurement units, and force-sensitive resistor sensors to capture precise embroidery actions. An extended Kalman filter is employed for robust multisensor data fusion and accurate motion trajectory estimation. The system was developed on the basis of a combined graph convolutional network–long short-term memory model, which showed a 95% accuracy in recognizing diverse needlework techniques by effectively capturing both spatial and temporal features of embroidery movements. This real-time recognition ability in an immersive augmented reality interface provides learners with dynamic visual guidance, step-by-step instructions, and performance feedback. The system overcomes the limitations of traditional preservation methods and offers a scalable, interactive, and effective platform for the transmission and documentation of traditional craftsmanship.

1. Introduction

As a vital component of China's intangible cultural heritage, traditional embroidery embodies centuries of artisanal mastery. However, its preservation has relied on master–apprentice oral instruction passed down from master to apprentice, leaving it vulnerable to decline as older-generation artisans retire. With the rapid advancement of augmented reality (AR) and human–computer interaction technologies, digital preservation of traditional craftsmanship becomes feasible. By integrating depth cameras with multisource sensors, AR systems provide real-time guidance for learners to practice complex needlework skills while digitally recording their operational processes, thereby ensuring interactive and immersive learning.

Researchers have paid attention to static pattern recognition or cultural exhibit displays, but not to the integration of multisensor action capture, graph model recognition, or AR interactive visualization into embroidery pedagogy. Therefore, we designed a system encompassing data

[*]Corresponding author: e-mail: zhaomingyu2794@dingtalk.com
https://doi.org/10.18494/SAM5799

perception, recognition inference, and visual feedback for the effective learning and preservation of traditional embroidery skills. The system captures embroidery actions through Kinect, Leap Motion, inertial measurement unit (IMU), and force-sensitive resistor (FSR) sensors and recognizes needle technique using the combined graph convolutional networks (GCNs) and long short-term memory (LSTM) model. The recognition results are visualized through real-time AR rendering. The system ensures a complete closed loop of action capture–recognition–feedback.(1) The developed system employs the Kalman filter with GCN + LSTM to capture embroidery motions. With a dynamically adjustable AR interface and real-time responsiveness based on recognition ability, the accuracy in needle technique recognition reached 95% in the experiment, demonstrating the model's effectiveness across diverse scenarios.

2. Related Works

Traditional embroidery craftsmanship is a vital part of intangible cultural heritage, yet its preservation and transmission face significant challenges, including an aging generation of artisans, the intricate nature of skill acquisition, and limited audience engagement. In this study, we aim to develop a system that integrates AR and deep learning for the digital and interactive preservation of traditional skills. Therefore, we conducted a literature review and analyzed previous research results in AR and sensor technologies in action recognition.

2.1 Embroidery skills

AI has proven effective in traditional textile pattern recognition. Shen embroidery, originating from Nantong in Jiangsu Province, China, is a representative form of intangible cultural heritage characterized by intricate needlework, vivid color contrasts, and motifs inspired by nature and daily life. Its patterns often emphasize fine detail and layered textures, making them both aesthetically distinctive and technically challenging to recognize computationally (Fig. 1). Shen embroidery is a specialized, modern branch of Suzhou embroidery, which is an ancient, broad tradition (over 2000 years old). A variant model of MobileNetV1 was used to recognize Shen embroidery patterns, achieving a recognition accuracy of 98.45%.(2) However, most studies have focused on recognizing static images, overlooking the dynamic movements and complex trajectories involved in the embroidery process.(3) The handcrafted processes in high-end fashion can be recorded through the motion capture ability of recent technologies for posture analysis. Considering the existing research results, we developed a method to recognize needle technique and provide prompt time feedback by integrating AR.

2.2 Gesture recognition on AR interface

AR has been widely applied in assembly guidance and manual training, enhancing the efficiency of complex task execution through overlaid guidance. Vision-based bare-hand AR has attracted attention owing to its natural interaction, affordability, and freedom from wearable devices or markers. However, achieving high precision in gesture recognition remains a

(a) (b)

Fig. 1. (Color online) Examples of Shen embroidery: (a) from https://www.suembroidery.com/chinese-silk-embroidery-blog/suzhou-embroidery-a-legacy-created-by-women and (b) from https://hand-su-embroidery.myshopify.com/products/delicate-hand-embroidered-suzhou-embroidery-art-chinese-mountains-landscape-50cm.

challenge. Recently, Leap Motion and Kinect have been used for stable and accurate finger-tracking in real-time interactive systems. Leap Motion and Kinect are sensor systems designed for motion tracking and human–computer interaction. Leap Motion uses infrared cameras to capture fine hand and finger movements with high precision, whereas Kinect employs depth sensing to track full-body motion in three dimensions. Leap Motion is well suited for detailed gesture recognition, whereas Kinect provides broader spatial coverage, making both valuable in applications such as virtual reality, robotics, and rehabilitation.(4) Such systems showed recognition rates of up to 88.2% and a latency of below 35 ms by integrating visual technologies such as a single-shot multibox detector (SSD) and continuously adaptive mean shift (CAMShift). Such previous developments can be implemented in embroidery action recognition through sensor data fusion.(5)

2.3 AR feedback in intangible cultural heritage education

AR systems used in cultural education enhance interactivity and visualization, providing immersive learning experiences. For example, real-time gesture feedback and AR-overlaid instructions are employed in embroidery training and related skill learning.(6) We employed real-time gesture feedback and AR-overlaid instructions to enhance the learning experience through virtual guidance trajectories, action indicators, and dynamic feedback.(7)

3. System Architecture and Fusion Algorithm

3.1 System architecture

In embroidery practice, learners wear AR glasses. A Kinect camera positioned in front of a learner captures full-body skeletal posture. A Leap Motion device mounted on the workstation tracks hand and fingertip movements,(8) whereas an IMU on the wrist and forearm enhances the recognition of hand rotation and posture.[9] FSR sensors embedded beneath the embroidery needle holder or fabric measure the threading force.[10] These sensors wirelessly transmit collected data to the computing layer for data fusion and processing. The system provides real-time visual feedback through the AR glasses, action guidance, practice steps, and performance metrics in immersive and interactive learning. The architecture of the developed system is illustrated in Fig. 2.

3.1.1 Perception layer

In the sensing layer, Kinect or RealSense captures upper-body skeletal posture, whereas Leap Motion tracks finger gestures. IMUs positioned on the wrist or elbow detect rapid rotational changes and compensate for drifts, ensuring precise motion tracking. FSR sensors positioned beneath the embroidery fabric or at the needle tip measure pressure during embroidery and tactile intensity during threading and thread pulling.

3.1.2 Computing layer

The sensor data fusion module is the most important component of the developed system. For temporal and spatial synchronization, an extended Kalman Filter (EKF) integrates Kinect positional data with IMU inputs to track smoothed six-degree-of-freedom motion trajectories. This method is widely used in AR and robotics, as it effectively minimizes jitter and frame loss. The system presents a skeletal graph, where key joints serve as nodes and limb connections form edges, with pressure measured by FSR sensors incorporated as additional node attributes. GCN extracts spatial features, whereas LSTM captures temporal dynamics. The outcomes include needlework categories, confidence scores, and trajectory evaluations to provide real-time feedback in the presentation layer.

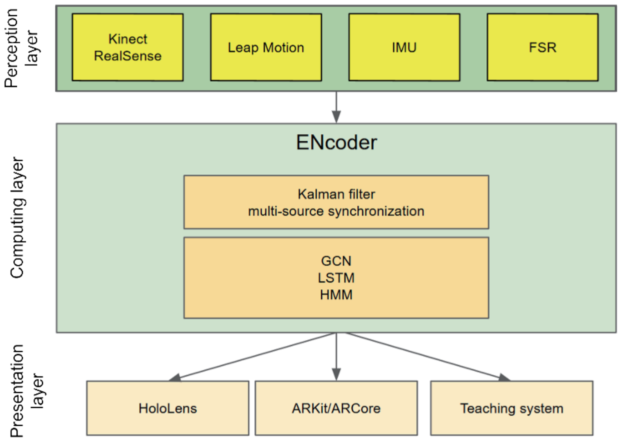

Fig. 2. (Color online) Architecture of system developed in this study. (HMM: hidden Markov model).

3.1.3 Presentation layer

The AR interface supports multiple devices, including head-mounted AR glasses (e.g., HoloLens) for immersive guidance, mobile apps for portable viewing, and desktop terminals for instructional evaluation. The display presents recognition results in real time, including backstitches, and overlays optimal trajectory paths or highlights needle exit points. After completing a stitch, numerical scores are provided as feedback, ensuring a seamless learning experience. With an update frequency of at least 30 frames per second (FPS), feedback is provided through gesture and voice. Virtual guidance is provided to the physical workspace, maintaining natural and intuitive interaction. Figure 2 illustrates the AR interface for guiding fishbone stitch operations, where the system segments the embroidery area into step zones (2, 3, and 4) (Fig. 3). Green virtual stitch trajectories are overlaid onto the physical fabric, pinpointing needle exit points and directional guidance, enabling learners to precisely follow operational paths.

3.2 Fusion algorithm

In the system, the accurate recognition of embroidery actions relies on data fusion and status estimation from Kinect, IMU, Leap Motion, and FSR. For effective motion trajectory processing, EKF is used to fuse multisensor data and enhance the system's robustness and accuracy for dynamic needlework actions.(11–13) The estimation algorithm with EKF fuses Kinect and IMU data as a link between the sensing and computing layers for smooth pose estimation and trajectory restoration in embroidery actions (Fig. 4).

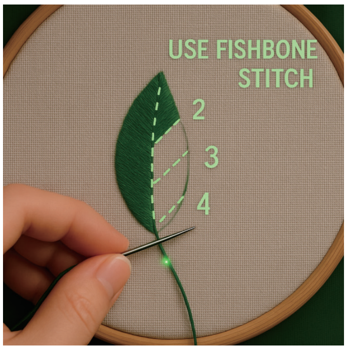

Fig. 3. (Color online) Example of teaching guide of developed system.

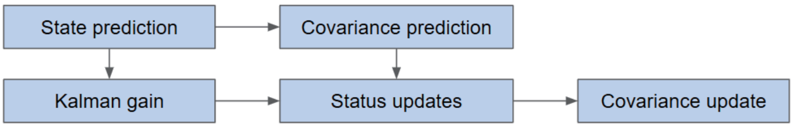

Fig. 4. (Color online) Flowchart of EKF's status estimation algorithm.

3.2.1 System status modeling

The status vector of the system is defined to represent the position and velocity of the embroidery action at each timestep x_k.

$$x_k = \begin{bmatrix} x \\ y \\ z \\ v_x \\ v_y \\ v_z \end{bmatrix} \tag{1}$$

Here, x, y, and z denote spatial position coordinates of the needle tip or hand joint positions, and v_x, v_y, and v_z present velocity components on the corresponding coordinate axes.

The system control input vector derived from IMU acceleration data is defined as u_k,

$$u_k = \begin{bmatrix} a_x \\ a_y \\ a_z \end{bmatrix}, \tag{2}$$

where a_x, a_y, and a_z denote acceleration components along the x-, y-, and z-axes in the IMU, reflecting wrist or arm action changes. The state transition equation of the prediction model is expressed as

$$x_k = Ax_{k-1} + Bu_k + w_k, \tag{3}$$

where A is the state transition matrix describing the kinematic evolution from timestep k–1 to k, B is the control input matrix mapping IMU data to the state space, and w_k represents process noise, assumed to follow a Gaussian distribution $w_k \sim N(0,Q)$. Then, the specific matrix form is defined as

$$A = \begin{bmatrix} 1 & 0 & 0 & \Delta t & 0 & 0 \\ 0 & 1 & 0 & 0 & \Delta t & 0 \\ 0 & 0 & 1 & 0 & 0 & \Delta t \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}, B = \begin{bmatrix} \frac{\Delta t^2}{2} & 0 & 0 \\ 0 & \frac{\Delta t^2}{2} & 0 \\ 0 & 0 & \frac{\Delta t^2}{2} \\ \Delta t & 0 & 0 \\ 0 & \Delta t & 0 \\ 0 & 0 & \Delta t \end{bmatrix}, \tag{4}$$

where Δt is the time interval between two consecutive samples.

3.2.2 Observation equation

The observation equation is defined as

$$z_k = Hx_k + v_k, \tag{5}$$

where z_k is the observed value (e.g., spatial positions directly measured by Kinect), H is the observation matrix that maps the state space to the observation space, and v_k is the observation noise, assumed to follow the distribution $v_k \sim N(0,R)$. The observation matrix H is defined as

$$H = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \end{bmatrix}. \tag{6}$$

3.2.3 EKF recursive update process

The computational workflow of EKF consists of the prediction and update phases. The prediction is conducted using Eqs. (7) and (8).

$$x_{k|k-1} = Ax_{k-1|k-1} + Bu_k \tag{7}$$

$$P_{k|k-1} = AP_{k-1|k-1}A^T + Q \tag{8}$$

Here, $x_{k|k\text{-}1}$ is the prior estimate of the state at time k and $P_{k|k\text{-}1}$ is the prior error covariance matrix, reflecting the uncertainty in the estimate.

On the basis of the equations, the Kalman gain is computed and updated as

$$K_k = P_{k|k-1}H^T\left(HP_{k|k-1}H^T + R\right)^{-1}. \tag{9}$$

State estimation is updated on the basis of measurements as follows.

$$x_{k|k} = x_{k|k-1} + K_k\left(z_k - Hx_{k|k-1}\right) \tag{10}$$

An error covariance matrix is updated as

$$P_{k|k} = \left(I - K_kH\right)P_{k|k-1}, \tag{11}$$

where K_k denotes the Kalman gain, representing the weight of observational corrections to the state estimate, $x_{k|k}$ is the posterior estimate at timestep k, $P_{k|k}$ is the posterior error covariance matrix describing the distribution of estimation errors, and I is the identity matrix.

The EKF algorithm of the developed system effectively fuses data from visual and inertial sensors for the real-time smoothed estimation and precise trajectory reconstruction of embroidery actions. The algorithm ensures robust real-time performance and high-precision feedback for AR embroidery education for digital heritage preservation.

4. Model Structure and Algorithm Design

EKF ensures the smoothing and precise estimation of needlework action trajectories. However, fused trajectory data alone is insufficient to solve the potential problems in automated recognition and real-time feedback for embroidery action classification.(14) Therefore, we have developed a needlework action recognition model by integrating GCN and LSTM for high-precision action classification and performance evaluation, enabling robust AR interactive feedback.

4.1 Action recognition model

We employed a joint model architecture to integrate spatial and temporal features for motion recognition in embroidery tasks. The system's recognition module identifies diverse needlework skills on the basis of fused multisensor action data, modeling the entire human–needle–fabric process in a spatiotemporal graph structure. In each temporal frame, graph nodes represent arm joints such as the wrist, elbow, and fingertips, with edges connecting these physical joints. Virtual nodes are introduced to represent the needle tip and fabric reference points, and each node incorporates feature vectors of 3D coordinates, pressure data, and angular variations.(15–17)

To capture the exact postural context, the graph was created by including head, shoulder, knee, and pelvic joints, which are a global reference for stability. Although embroidery techniques are executed primarily with the hands, traditional practitioners maintain a specific seated posture that stabilizes the upper body. Including lower-body joints allows GCN to normalize upper-body trajectories against a fixed base, reducing noise caused by seating shifts or overall body movement. The model's attention mechanism emphasizes the high-frequency movements of the hands while using the lower joints as static anchors for coordinate normalization. The detailed skeletal graph structure is depicted in Fig. 5.

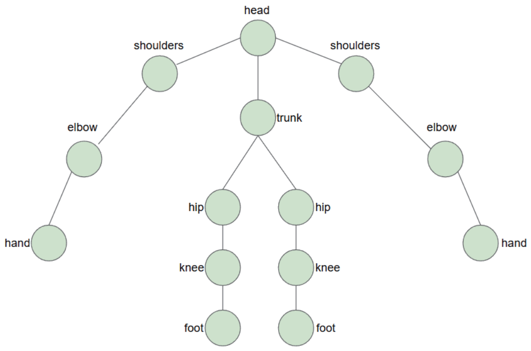

Fig. 5. (Color online) Skeletal graph structure of action recognition model.

We defined the skeletal graph structure for embroidery action capture, in which key nodes of the head, shoulder, elbow, wrist, and fingertips constitute the node set *V*, and skeletal connections form the edge set *E*. This is represented as

$$G = (V, E), V = \{v_1, v_2, \ldots, v_n\}, E \subseteq V, \tag{12}$$

where *V* denotes skeletal key points of the head, shoulder, hand, and knee, and *E* represents naturally connected skeletal edges. The adjacency matrix *A* explicitly describes connectivity and relationships between nodes, effectively capturing the spatial dependencies of embroidery actions through the graph structure. The adjacency matrix is defined as

$$A_{ij} = \begin{cases} 1 & \text{if } v_i \text{ is connected to } v_j, \\ 0 & \text{otherwise.} \end{cases} \tag{13}$$

Typically, the adjacency matrix is normalized to optimize network training, denoted as $\hat{A} = A + I$.

4.1.1 GCN

We used graph convolution to extract the spatial features of the nodes, with the propagation equation

$$H^{(l+1)} = \sigma\left(\hat{D}^{-1/2} \hat{A} D^{-1/2} H^{(l)} W^{(l)}\right), \tag{14}$$

where $H^{(l)}$ represents node features at the *l*-th layer, $\hat{D}$ is the degree matrix of the adjacency matrix, $W^{(l)}$ denotes the learnable weight matrix, and σ is the nonlinear activation function. The final output $H^{(L)}$ is fed into a softmax classification layer to produce action labels such as threading and coiling.

4.1.2 Model training and loss function

The GCN + LSTM model was trained using the multimodal spatiotemporal dataset collected during the experiment. The training data comprised 4800 labeled sequences of embroidery actions, with each sample including synchronized 3D skeletal coordinates (Kinect), fine-grained hand gestures (Leap Motion), rotational dynamics (IMU), and tactile pressure values (FSR). The dataset consists of 4,800 labeled action samples covering four primary techniques: threading, coiling, backstitching, and cross stitching.(18) To train the GCN model, the cross-entropy loss function is defined as the optimization objective.

$$LL = -\sum\nolimits_{i=1}^{C} y_i \log \hat{y}_i \tag{15}$$

Here, y_i denotes the real action class label, $\hat{y}_i$ represents the predicted probability of the corresponding class by the GCN network, and C is the total number of action categories for classification. Optimizing this loss function enables the model to better identify different embroidery actions, thereby improving overall classification accuracy.

4.1.3 Input feature

In the input process, the initial input features of each node are defined as $h_i^{(0)}$, which are generally composed of 3D positional coordinates, velocity values, local joint angles, fingertip pressure, and posture angles captured by IMU. These features collectively describe the spatial dynamic information during the embroidery process, ensuring the accurate extraction and recognition of actions of the subsequent network.

4.2 GCN + LSTM model

To sufficiently capture both spatial and temporal information simultaneously, we developed an architecture by integrating GCN and LSTM.(19,20) The structure with detailed workflow and algorithmic steps of the model is illustrated in Fig. 6.

4.2.1 Input layer

The input layer consists of GCN and LSTM branches. In the GCN branch, the input is a graph structure composed of skeletal key points per frame (joint nodes + connecting edges), formatted as graph convolution input.

$$X_t \in R^{V \times d} \tag{16}$$

Here, V is the number of nodes, and d denotes the feature dimension per node including 3D coordinates and velocity.

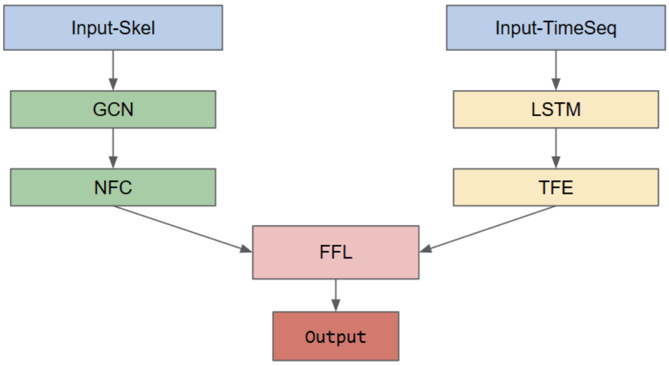

Fig. 6. (Color online) GCN+LSTM model architecture.

In the LSTM branch, the input is a continuous-frame sequence of full-body or localized action, formatted as

$$SS = \left[x_1, x_2, \ldots, x_T\right], x_t \in R^n , \tag{17}$$

where T represents the sequence length and n indicates the feature dimension per frame of global posture, hand posture, velocity, and acceleration.

4.2.2 Coding layer

The GCN branch performs graph convolution operations on the skeletal node graph, aggregating spatial relationships between nodes and their adjacent counterparts. The GCN branch models the local spatial topological structures in embroidery actions. The LSTM branch conducts temporal modeling on the input sequential action frames, capturing the dynamic temporal evolution of actions during embroidery to encode the time-dependent features of hand gestures. Through collaborative operation, the model leverages complementary strengths in spatial structural perception and temporal dynamic modeling, providing deep feature support for subsequent high-precision embroidery action recognition.

4.2.3 Feature fusion layer

In the feature fusion layer, a weighted concatenation method is employed to jointly encode the output features from both branches.

$$h_{fusion} = \alpha \cdot h_{GCN} + \left(1 - \alpha\right) \cdot h_{LSTM} \tag{18}$$

Here, h_{GCN} denotes the spatial features output by the GCN branch, h_{LSTM} represents the temporal features from the LSTM branch, and α is a fusion weighting factor within the range of (0,1), balancing the contributions of spatial and temporal features. The value is adaptively learned through network training.

4.2.4 Output layer

The fused features are passed through a fully connected layer and then classified by a softmax classifier to generate the final predicted probabilities for embroidery action categories, such as Panjin stitch, cross stitch, and backstitch.

$$\hat{y} = \mathrm{softmax}(Wh_{fusion} + b) \tag{19}$$

Here, W denotes the weight parameters of the fully connected layer, b is the bias term, and $\hat{y}$ represents the final output probability vector for action categories, including specific embroidery techniques such as Panjin stitch, cross stitch, and backstitch.

In the GCN + LSTM model, sensors first capture dynamic action data of the hand and needle tip during embroidery. Then, the model uses a dual-channel input that comprises skeletal node graph structures and continuous action sequences. In the model, the GCN branch employs graph convolution mechanisms to extract spatial topological relationships between nodes and captures coordinated variations in fingertip postures and joint interactions during embroidery. The LSTM branch learns the temporal evolutionary patterns of embroidery actions to uncover dynamic patterns in needlework. The output features from the branches are fused in a weighted concatenation layer and classified by a Softmax classifier to output embroidery action categories and labels.

Recognition accuracy was defined as the ratio of correctly predicted action labels to the total number of action sequences in the test set. Latency was measured as the mean processing time per frame on a Jetson Nano edge computing device, encompassing data fusion (EKF), model inference, and AR rendering. Model size was quantified by the number of trainable parameters and the final binary file size after 8-bit quantization.

In the experiment, the GCN + LSTM model showed superior performance, which was evaluated by using recognition accuracy, latency, and model size. The overall recognition accuracy was 95%, higher than that of a single model of LSTM (89%) and GCN (91%). The combined model was more effective in addressing the recognition challenges posed by complex action trajectories and strong spatiotemporal coupling in embroidery techniques. In terms of inference latency, LSTM exhibited the fastest response, making it appropriate for embedded scenarios with stringent real-time requirements. While the GCN + LSTM model showed a slightly higher latency, its precision advantages in teaching feedback and skill assessment contexts were substantial, especially with instructional and evaluative capabilities. The CNN + LSTM model, as an intermediate solution, achieved slightly lower accuracy than the GCN + LSTM model and demonstrated spatial feature extraction ability inferior to that of the GCN model (Fig. 7).

Considering deployment costs and device compatibility, although the GCN + LSTM model is larger than other models, it ensures lightweight operation on edge computing devices such as Jetson Nano and AR headsets, as the model adopts model pruning and parameter quantization. The GCN + LSTM model demonstrates strong practicality and scalability in AR embroidery

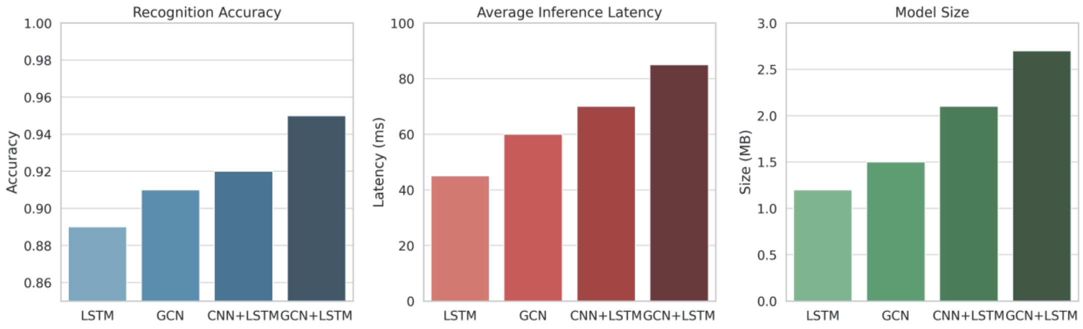

Fig. 7. (Color online) Results of performances of different models in this study.

interactive teaching. It allows for flexible model architecture selection tailored to application requirements, facilitating collaborative adaptation across diverse scenarios, such as exhibition, instruction, and evaluation. This model serves as an essential system for the preservation of intangible cultural heritage craftsmanship.

5. Results and Discussion

5.1 Experiment

We recruited 24 participants (all females aged 19–45 years old). The participants included eight professional embroidery practitioners with more than five years of experience and sixteen college students with little to no prior embroidery training. The experiment was conducted in a controlled environment, with each session lasting approximately 60 min. The experiment was conducted by the following procedure.

- Calibration: All sensors were calibrated to match the participant's physical dimensions and seating posture.
- Training: The participants performed four fundamental Shen embroidery actions: threading, coiling, backstitching, and cross stitching.
- Data acquisition: Each action was repeated 50 times by each participant, resulting in a dataset of 4800 action samples.

The college students completed a specific embroidery pattern with real-time AR guidance and trajectory monitoring. The experimental hardware comprised multiple sensors to capture the multifaceted nature of Shen embroidery. A Kinect V2 was positioned 1.2 m in front of each participant to track upper-body skeletal posture. In addition, a Leap Motion controller was mounted on the embroidery workstation to record high-precision hand and finger movements. The participants wore a six-axis IMU on their dominant wrist to measure rapid rotational data. Four FSRs were embedded beneath the embroidery fabric and attached to the needle holder to detect threading and pulling forces during stitching. Epson BT-35E AR glasses were used to provide an immersive visual interface and deliver real-time feedback. As embroidery materials, standard silk fabric and silk threads were used to present Shen embroidery techniques.

5.2 Results

To validate the adaptability of the GCN + LSTM model, we evaluated the model's recognition accuracy, system responsiveness, and deployment compatibility in embroidery tasks. For the evaluation of the model performance, we constructed matrices for action recognition, confusion matrices, and scenario-specific recommendations, and measured the frame rate impact on stability.

5.2.1 Embroidery action recognition

A confusion matrix was constructed by comparing the ground-truth labels (annotated by master practitioners during data collection) with the model predictions (the class label with the highest probability output by the Softmax classifier) (Tables 1 and 2). The results enable the identification of misconceptions or overlapping features between similar techniques, such as coiling and backstitching. The confusion matrix revealed variations in recognition accuracy in different actions. Threading and cross stitching showed almost no misclassification, whereas coiling and backstitching exhibited partial overlap, highlighting the need to detect rotational and pull-back motions more accurately. The experimental result showed that the GCN + LSTM model presented the best discriminative capability for needlework actions and strong distinctiveness in the cross stitching phase.

Figure 8 shows how recognition accuracy changes with different FPS. This result demonstrated that higher frame rates improved the system's ability to recognize embroidery actions. With increasing FPS, the model captures finer temporal details of hand and needle movements, reducing ambiguity and misclassification. The improvement is most noticeable between 20 and 30 FPS, whereas gains from 30 to 60 FPS are smaller but still meaningful. This indicates that while moderate FPS (30) already provides strong recognition, higher FPS (60)

Table 1
Confusion matrix.

Actual result	Predicted result			
	Threading	Coiling	Backstitching	Cross stitching
Threading	1152	48	0	0
Coiling	0	1080	120	0
Backstitching	0	0	1104	96
Cross stitching	0	0	0	1176

Table 2
Results of action recognition.

Action	Prediction accuracy (%)	Common misconceptions
Threading	96	Occasional misclassification as coiling
Coiling	90	Confused with backstitching
Backstitching	92	Occasional misclassification as cross stitching
Cross stitching	98	Almost no misidentification

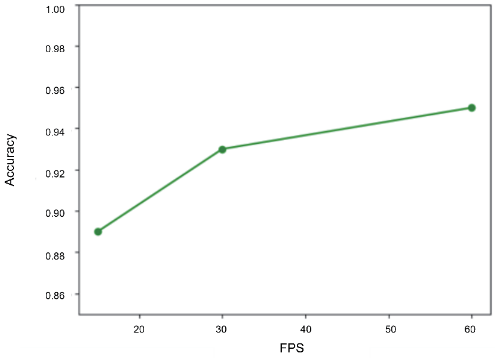

Fig. 8. (Color online) Performance evaluation of CGN + LSTM model using relationship between prediction accuracy and FPS.

offers optimal accuracy for applications requiring precise tracking, such as training novices or documenting expert techniques.

5.2.2 Effect of system frame rate on recognition ability

Given the performance differences of different models, we tested the system's average recognition accuracy on the same action set at 15, 30, and 60 FPS and presented the results in Table 3. To determine recognition accuracy at various frame rates, we performed a controlled test using the same validation dataset. The original 60 FPS sensor data was downsampled to 15 and 30 FPS. The model's accuracy was then recalculated for each frequency, revealing that 30 FPS provides the optimal balance between recognition precision and computing resource efficiency for real-time AR feedback. When the frame rate increased from 15 to 30 FPS, the accuracy improved significantly. From 30 to 60 FPS, the accuracy was stabilized. On the basis of recognition ability and computing resource balance, a minimum sampling frequency of 30 FPS was found to be appropriate for the accurate recognition of action initiation and transitional frames of the model.

5.2.3 Action distribution

To validate the model's classification accuracy in high-dimensional action feature spaces, we used the t-distributed stochastic neighbor embedding (t-SNE) dimensionality reduction algorithm to visualize multimodal features, such as skeletal poses and FSR, and project them onto a 2D plane. Figure 9 presents the distribution of features extracted by the GCN + LSTM

Table 3
Effect of different frame rates on recognition ability.

Frame rate (FPS)	Accuracy
15	89%
30	93%
60	95%

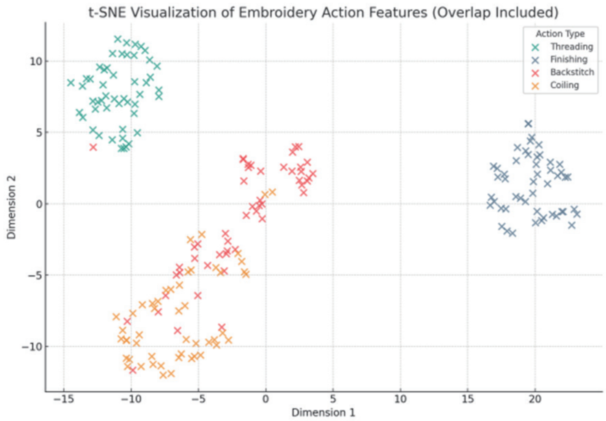

Fig. 9. (Color online) Embroidery action distribution map of GCN + LSTM model in action feature space.

Table 4
Scenario-adaptive model recommendation matrix.

Scenario	Model recommendation	Description
Museum AR display	GCN	Stable recognition with low resource consumption, suitable for embedded devices
Intangible cultural heritage teaching platform	GCN + LSTM	Dual-channel model achieves high recognition accuracy, ideal for training feedback
Fast mobile interaction	LSTM	Low inference latency and easy deployment, compatible with APPs and Pads
Professional embroidery training system	GCN+LSTM	Capable of modeling complex actions and supporting scoring feedback, applicable to educational scenarios

model in the action feature space. Four typical needlework actions were clustered, reflecting the model's effectiveness in semantic action recognition. In the figure, four needlework actions are marked in different colors: green for threading, blue for coiling, red for backstitching, and orange for cross stitching. The clusters confirm the model's ability to categorize actions with boundaries, especially between threading and cross stitching, highlighting the separability of actions. The partial overlap of coiling and backstitching suggests that recognition accuracy needs to be improved by enhancing dynamic constraint modeling.

5.3 Model recommendation

Considering the varying demands in diverse applications and the recognition accuracy, response speed, and deployment resources of the models, we compared different models as shown in Table 4. The table can be used as a reference to balance recognition precision, feedback latency, and device capabilities for institutions, organizations, and mobile developers and provide a closed-loop experience in visualizable, learnable, and evaluable intangible cultural heritage craftsmanship and education. GCN has been shown to perform well in structured environments such as museum AR displays, where stable recognition and low resource consumption are critical for embedded systems.(21) LSTM is preferred in fast mobile interaction scenarios owing to its low inference latency and ease of deployment on lightweight devices.(22) For educational platforms and professional embroidery training systems, combining GCN and LSTM enables the dual-channel modeling of spatial and temporal features, enhancing recognition accuracy and enabling dynamic feedback and scoring.(6)

6. Conclusions

We developed and validated a digital inheritance system for traditional embroidery skills, offering a robust solution for the preservation of intangible cultural heritages. By integrating multisource data from Kinect for skeletal posture to FSR sensors for threading force and employing an EKF, the system ensures highly accurate status estimation and motion trajectory reconstruction of complex needlework. The GCN-LSTM model presented a remarkable 95%

recognition accuracy for embroidery actions, significantly outperforming single models such as CNN and LSTM. This superior performance is attributed to the model's ability to concurrently analyze spatial relationships within skeletal movements using GCN and temporal dynamics of actions using LSTM, thus enhancing recognition robustness and minimizing challenges such as occlusion. The system's real-time AR feedback mechanism provides learners with immersive and personalized learning, offering dynamic demonstrations, step-by-step guidance, and immediate performance evaluations. This transcends the inherent spatiotemporal constraints of traditional master–apprentice instruction and ensures the digital documentation of operational processes, mitigating the risk of the loss of skill as older generations retire. The developed system with instruction, exhibition, and evaluation enables a standardized, platform-based, and sustainable approach to intangible cultural heritage preservation, advocating for national-level coordination in policy, funding, and interdisciplinary talent development. It is necessary to integrate haptic gloves for force feedback, establish a cloud platform for evaluation, and support multilingual and cross-cultural databases to disseminate and exchange intangible cultural heritage craftsmanship.

To advance digital learning systems for traditional embroidery using sensors, AR, and deep learning, national-level coordination and standardized frameworks are required. A lifecycle standard covering data acquisition, modeling, interaction, and presentation should be established to ensure interoperability and align with international charters on digital heritage. A unified platform for 3D action and multimodal cultural resources supports rapid data upload, retrieval, visualization, and integration with AI models, software development kits, and open APIs to enable research, teaching, and secondary development. Sustained government support is essential, including funding, tax incentives, and investment in AR teaching products for schools and cultural centers. Talent development requires interdisciplinary training programs in higher education, dual mentorship between heritage masters and technical teams, and initiatives such as workshops and competitions to enhance digital literacy. Legal frameworks must define ownership, establish digital asset registration, and enforce copyright protection to safeguard intangible cultural heritage against misuse and ensure fair benefit distribution.

Acknowledgments

This research was supported by the Basic Scientific Research Funds for Provincial Universities in Heilongjiang for the project "Research on Brand Building of Heilongjiang Intangible Cultural Heritage Based on Brand Narrative Theory" (Grant No. 2025-KYYWF-E015). This research was supported by the Horizontal Research Project of Harbin Finance University, "Research on Intelligent Short Video Generation and Brand Communication Strategy of Intangible Cultural Heritage Traditional Embroidery Based on Deep Learning".

References

1 R. G. Boboc, E. Băutu, F, Gîrbacia, N. Popovici, and D.-M. Popovici: Appl. Sci. **12** (2022) 9859. https://doi.org/10.3390/app12199859

2 J. Zhu and C. Zhu: Sci. Rep. **14** (2024) 9574. https://doi.org/10.1038/s41598-024-60121-7

3 M. Eswaran, V. V. S. S. Prasad, and M. Hymavathi: J. Manuf. Syst. **72** (2024) 104. https://doi.org/10.1016/j.jmsy.2023.11.002
4 T. Guzsvinecz, V. Szucs, and C. Sik-Lanyi: Sensors **19** (2019) 1072. https://doi.org/10.3390/s19051072
5 W. Fang, L. Zheng. and X. Wu: Comput. Ind. **92** (2017) 91. https://doi.org/10.1016/j.compind.2017.06.002.
6 Q. Yu, X. Tao, and J. Wang: Sustainability **17** (2025) 7657. https://doi.org/10.3390/su17177657
7 R. Mao: Proc. IEEE 6th Int. Seminar Artificial Intelligence, Networking and Information Technology (AINIT, 2025) 1761. https://doi.org/10.1109/AINIT65432.2025.11035726
8 J. Guo, H. Liu, X. Li, D. Xu, and Y. Zhang: Appl. Sci. **11** (2021) 8641. https://doi.org/10.3390/app11188641
9 S. Zhang, Y. Li, S. Zhang, F. Shahabi, S. Xia, Y. Deng, and N. Alshurafa: Sensors **22** (2022) 1476. https://doi.org/10.3390/s22041476
10 J. Dong J, Y. Gao, H. J. Lee, H. Zhou, Y. Yao, Z. Fang, and B. Huang: Appl. Sci. **10** (2020) 1482. https://doi.org/10.3390/app10041482
11 C. Li, A. Fahmy, and J. Sienz: Sensors **19** (2019) 4586. https://doi.org/10.3390/s19204586
12 L. Tanzi, P. Piazzolla, S. Moos, and E. Vezzetti: Int. J. Interact. Des. Manuf. **17** (2023) 103. https://doi.org/10.1007/s12008-022-01107-5
13 N. Menon, E. S. L. Vasquez, H. Curran, S. D. Koninck, and L. Devendorf: Proc. 2023 ACM Int. Joint Conf. Pervasive and Ubiquitous Computing and the 2023 ACM Int. Symp. Wearable Computing (ACM, 2023) 310. https://doi.org/10.1145/3594739.3610786
14 Y. Miky, Y. Alshawabkeh, and A. Baik: Herit. Sci. **12** (2024) 255. https://doi.org/10.1186/s40494-024-01382-3
15 M. Noor-ul-Huda, H. Ahmad, A. Banjar, A. O. Alzahrani, I. Ahmad, M. S. Naeem: Heliyon **10** (2024) e26466. https://doi.org/10.1016/j.heliyon.2024.e26466
16 A. S. Nunes, İ. Y. Potter, R. K. Mishra, J. Casado, N. Dana, A. Geronimo, C. G. Tarolli, R. B. Schneider, E. R. Dorsey, J. L. Adams, and A. Vaziri: Commun. Med. **5** (2025) 50. https://doi.org/10.1038/s43856-025-00770-5
17 K.-B. Park, M. Kim, S. H. Choi, and J. Y. Lee: Robot. Comput.-Integr. Manuf. **63** (2020) 101887. https://doi.org/10.1016/j.rcim.2019.101887
18 X. Zhou: Int. J. Agric. Environ. Inf. Syst. **16** (2025) 1. https://doi.org/10.4018/IJAEIS.392030
19 C. Si, W. Chen, W. Wang, L. Wang, and T. Tan: arXiv 2019. https://doi.org/10.48550/arXiv.1902.09130
20 L. Wang, A. Xu, M. Wei, W. Zuo, and R. Li: J. Manuf. Syst. **73** (2024) 307. https://doi.org/10.1016/j.jmsy.2024.02.009
21 R. Pierdicca, M. Paolanti, S. Naspetti, S. Mandolesi, R. Zanoli, and E. Frontoni: J. Imaging **4** (2018) 101. https://doi.org/10.3390/jimaging4080101
22 X. Pan, G. Huang, Z. Zhang, J. Li, H. Bao, and G. Zhang: IEEE Trans. Vis. Comput. Graph. **30** (2024) 7354. https://doi.org/10.1109/TVCG.2024.3456152

About the Authors

Xuehong Zhao received her M.S. degree in industrial economics from Harbin University of Science and Technology in 2009. Since 2021, she has been an associate professor at Harbin Finance University, China. Her research interests include intangible cultural heritage communication, brand management, and artificial intelligence. (2009029@hrbfu.edu.cn)

Mingyu Zhao received his M.S. degree in software engineering from Northeastern University in 2015. Since 2011, he has served as an assistant researcher at Heilongjiang Academy of Science, focusing on deep learning algorithms and sensor technology. (zhaomingyu2794@dingtalk.com)

Hailing Wang received her M.S. degree in computer application technology in 2009 and her Ph.D. degree in computer application technology from Harbin Engineering University, P.R. China, in 2013. Since then, she has served as a lecturer at Heilongjiang University of Science and Technology, specializing in knowledge graphs, virtual reality, and software engineering.
(wanghailing@usth.edu.cn)