IIETA International Information and Engineering Technology Association

Traitement du Signal
Vol. 42, No. 6, December, 2025, pp. 3409-3421
Journal homepage: http://iieta.org/journals/ts

Trajectory Analysis of Learning Processes for Personalized Education: Deep Spatiotemporal Embedding and Discovery of Learning Behavior Patterns

Qiaojuan Lei

College of Pre-school Education, Xi'an University, Xi'an 710065, China

Corresponding Author Email: 18691550175@163.com

https://doi.org/10.18280/ts.420628

Received: 7 June 2025
Revised: 2 November 2025
Accepted: 15 November 2025
Available online: 31 December 2025

Keywords:
personalized education, learning process images, trajectory analysis, spatiotemporal embedding, behavior pattern discovery, Swin transformer, lightweight models

ABSTRACT

The in-depth development of personalized education urgently demands accurate capture of dynamic learning process trajectories. Existing methods show significant limitations in modeling spatiotemporal correlations in dual-phase learning images, decoding trajectory semantics, and generalizing behavior patterns, making it difficult to support process-oriented learning diagnostics effectively. To address this, we propose an end-to-end framework that integrates feature extraction, spatiotemporal embedding, trajectory analysis, and pattern discovery. We first use a parameter-shared dual-branch Swin Transformer V2 to extract multi-scale features from dual-phase learning images, enhanced by a multi-scale differential fusion module to emphasize trajectory changes. A spatiotemporal embedding mechanism maps features into high-dimensional trajectory vectors, and after reconstructing the full trajectory using Dynamic Time Warping (DTW), we apply an improved Density-Based Spatial Clustering of Applications with Noise (DBSCAN) algorithm to discover learning behavior patterns. A lightweight strategy and contrastive loss are introduced to balance model accuracy and efficiency. Experimental results demonstrate that the proposed spatiotemporal embedding representation outperforms mainstream embedding and traditional clustering methods, achieving a clustering purity of up to 0.92. Key modules collaborate synergistically, with the multi-scale differential fusion module playing a crucial role; its removal reduces the F1 score of pattern recognition by 15.6%. Contrastive loss reduces trajectory cluster overlap from 42% to 11%. Personalized intervention shows significant effects, reducing task completion time by 25.6% for randomly exploratory learners and increasing test scores by 13.2% for indecisive learners. This study shows that the proposed framework enables an end-to-end transformation from image features to interpretable behavior patterns, providing critical technical support for real-time interventions and learning path optimization in personalized education, and enriching interdisciplinary research paradigms in educational data mining.

1. INTRODUCTION

The deepening advancement of personalized education has prompted a shift in the educational assessment paradigm from a traditional results-oriented approach to a process-oriented one, with dynamic trajectories in the learning process becoming the core basis for analyzing learners' cognitive states and learning strategies [1, 2]. Learning process image trajectories, such as eye movement focus shifts and handwritten operation sequences, can intuitively map learners' attention distribution, thought progression, and decision-making processes, providing rich unstructured data support for precise learning condition diagnosis [3-5]. However, learning trajectories simultaneously possess the dual attributes of spatiotemporal continuity and semantic ambiguity: at the microscopic level, it is necessary to capture the temporal correlations of eye movement points and computational steps, and at the macroscopic level, it is necessary to interpret the semantic mapping of trajectory distribution and learning efficiency. Existing methods struggle to achieve the coordinated modeling of both [6, 7]. Meanwhile, the computational power constraints of edge devices in educational scenarios, such as smart tablets and classroom terminals [8, 9], impose dual demands for high-precision trajectory analysis and lightweight deployment, making traditional complex models unable to meet this practical need.

Existing research in the field of learning trajectory analysis mostly relies on structured data, such as answer time and click sequences [10, 11], and has insufficient exploration of the spatiotemporal features contained in image trajectories. A complete decoding process, from feature extraction to trajectory reconstruction to pattern induction, has not been established, leading to a disconnect between technological outputs and educational semantic interpretation. In the areas of dual-phase feature extraction and spatiotemporal embedding, the Transformer architecture, with its powerful temporal modeling ability, has shown significant advantages in dual-phase data processing tasks, but the application of parameter-sharing dual-branch architectures in educational trajectory analysis is rare [12-14]. Existing spatiotemporal embedding techniques mostly focus on general scenarios such as pedestrian trajectories and traffic flow, failing to adapt to the unique logical associations of learning trajectories, such as the semantic progression from problem analysis to solution, resulting in embedding vectors lacking interpretability in educational scenarios [15, 16]. In terms of behavior pattern discovery, traditional statistical clustering methods, such as K-Means, still dominate the education field [17-19], with low integration with deep learning feature extraction processes, making it difficult to capture the semantic association patterns implicit in the trajectories, thus limiting the educational practical value of the pattern discovery results.

The core gaps in existing research can be summarized in three aspects: First, the feature extraction of dual-phase learning images lacks trajectory-oriented design, and the feature representation and spatiotemporal continuity of trajectories have not formed effective binding, resulting in insufficient capture accuracy of key trajectory turning points, such as attention shifts; second, deep spatiotemporal embedding is disconnected from educational semantics, as existing embedding vectors only represent feature-level similarity and cannot map to trajectory segments with clear educational semantics, such as problem analysis and computation; third, the trajectory analysis and pattern discovery processes are disjointed, lacking an integrated flow from trajectory segment segmentation to pattern clustering, making it difficult for technological outputs to directly translate into educational decision support information. Based on this, this study constructs a technology framework that balances precision and efficiency through the collaborative design of parameter-sharing dual-branch architecture, multi-scale differential fusion, spatiotemporal embedding decoding, and pattern clustering, achieving an organic integration of technological innovation and educational value.

The core contributions of this study include two dimensions: theoretical and technical. On the theoretical level, we propose a three-level mapping mechanism of dual-phase features, spatiotemporal embedding, and trajectory semantics, establishing a deep representation framework for learning trajectories, filling the research gap in the semantic interpretation of spatiotemporal embedding in educational scenarios; we also build a behavior pattern discovery process of trajectory segmentation, clustering, and semantic annotation, achieving a precise transformation from technological features to educational semantics. On the technical level, we design a collaborative architecture of parameter-sharing dual-branch Swin Transformer V2 (STv2) and multi-scale differential fusion, enhancing the precision of spatiotemporal correlation modeling through cross-phase feature consistency constraints and multi-granularity feature aggregation; we introduce a combination strategy of feature dimensionality reduction and lightweight decoding, along with contrastive loss optimization to improve embedding performance, enabling high-precision trajectory analysis and real-time pattern recognition on edge devices.

The subsequent sections of this paper are arranged as follows: The second part details the technical specifics of the proposed integrated framework, including core module designs for feature extraction, spatiotemporal embedding, trajectory analysis, and pattern discovery; The third part verifies the effectiveness of the method through ablation experiments and baseline comparison experiments; The fourth part discusses the core value, limitations, and future directions of the research findings; Finally, the conclusion summarizes the entire paper.

2. METHODS

2.1 Problem definition

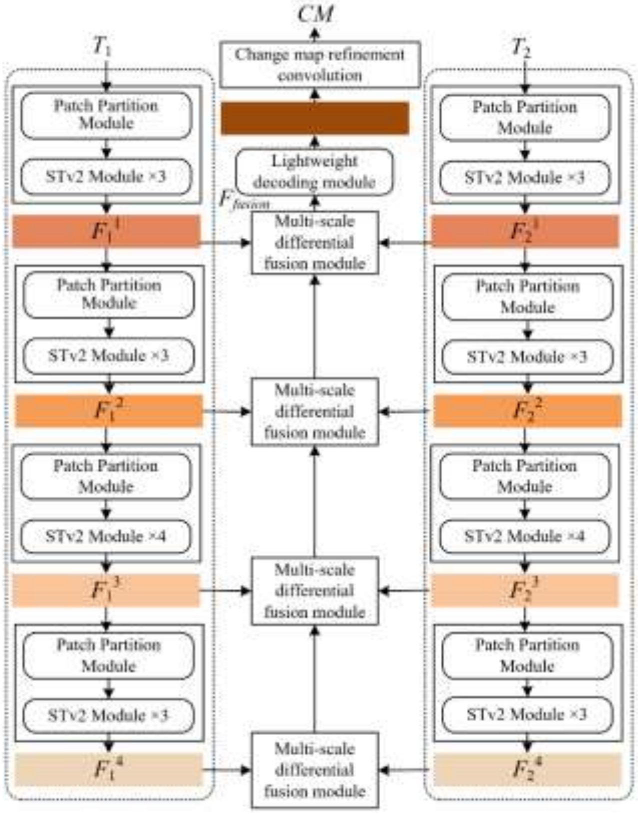


Figure 1. Model architecture for learning trajectory analysis based on parameter-sharing dual-branch STv2 + multi-scale differential fusion

This chapter aims to address the problem of trajectory analysis and behavior pattern discovery for dual-phase learning process images, providing clear mathematical definitions and semantic meanings for the input, intermediate representations, and output, setting clear objectives for the subsequent technical framework design. The input to this problem consists of two learning process image pairs captured at different moments T_1 and T_2: $I_1 \in \mathrm{R}^{H \times W \times 3}$ and $I_2 \in \mathrm{R}^{H \times W \times 3}$, where H and W represent the image height and width, respectively, and 3 is the number of RGB channels. The image contents are visualized encoded results of learning trajectories, such as eye movement focus and handwritten operations. Based on this input, the model first generates two key intermediate outputs: First, the deep spatiotemporal embedding vector of the learning trajectory $E \in \mathrm{R}^{N \times D}$, where N is the total number of trajectory sampling points, and D is the embedding dimension. The vector must encode both the spatiotemporal position information and semantic association features of the trajectory; Second, the trajectory segment set $S=\{S_1, S_2, \ldots, S_M\}$, where M is the number of segments, and each segment corresponds to a continuous trajectory segment with unified semantics. The final output of the problem includes two types of results: First, the learning behavior change map $CM \in \mathrm{R}^{H \times W \times 1}$, which represents the spatial distribution of learning behavior changes between the two moments at the pixel level; second, the learning behavior pattern label set $L=\{l_1,l_2,...,l_K\}$, where K is the number of pattern categories, and the labels must accurately correspond to typical learning behavior patterns with educational semantics. Figure 1 shows the complete model architecture for learning trajectory analysis based on parameter-sharing dual-branch STv2 + multi-scale differential fusion.

2.2 Parameter-sharing dual-branch Swin Transformer V2 encoder

The core function of the parameter-sharing dual-branch STv2 encoder is to perform multi-scale feature extraction and trajectory-oriented representation for dual-phase learning process images. Its design centers on ensuring feature consistency across the dual-phase and encoding multi-granularity trajectory information. To address the trajectory discontinuity issue caused by cross-phase feature space shifts in traditional dual-branch structures, the encoder adopts a connected network architecture, where both branches share the complete STv2 backbone network weights. This design enforces feature extraction of dual-phase images within a unified semantic space by binding the weights, thus providing the basis for subsequent temporal continuity modeling of trajectories, while also reducing model parameters, laying the foundation for lightweight deployment.

The encoder employs a four-stage progressive feature extraction process, which uses downsampling and module stacking to achieve multi-granularity encoding of trajectories from micro to macro levels. The input images are first converted into a sequence of image patches using the 4×4 Patch Partition operation, then processed in the first stage: Three STv2 modules are applied to the image patch sequence for feature encoding, generating scale 1 features with a resolution of H/4×W/4. These features focus on the spatial positions and morphological information of micro trajectories, such as eye movement points and pen tip landing points. After the first stage, subsequent stages perform 2x downsampling via Patch Merging operations, while adjusting the number of stacked STv2 modules based on the granularity of trajectory information: The second stage stacks three STv2 modules, generating scale 2 features with a resolution of H/8×W/8, encoding the associations of local continuous trajectory segments; The third stage stacks four STv2 modules, generating scale 3 features with a resolution of H/16×W/16, primarily modeling the logical associations of trajectories, such as from the problem stem to the options, or from the formula to the calculation region. The increased number of modules is due to the complexity of trajectory logical associations requiring deeper feature interactions; The fourth stage stacks three STv2 modules, generating scale 4 features with a resolution of H/32×W/32, encoding global trajectory distribution patterns, such as the spatial transition features between focused and scattered regions. The multi-scale feature pyramid formed after the four stages fully covers the microscopic morphology, local associations, and global distribution information of the trajectory.

2.2.1 Architecture design

The architecture design of the parameter-sharing dual-branch STv2 encoder is based on the core principles of "trajectory continuity modeling" and "multi-granularity information adaptation," achieving a balance between the precision and efficiency of feature extraction through structural constraints and process optimization. The design logic of the dual-branch connected structure stems from the core requirement of dual-phase trajectory analysis—the comparability of features across phases: If the two branches use independent weights, the feature expressions of the same trajectory pattern at different moments would be misaligned, leading to misjudgments of key trajectory turning points. The weight-sharing mechanism ensures that dual-phase images generate features under the same feature extraction rules, maintaining feature consistency for the same trajectory pattern and providing a reliable basis for subsequent trajectory temporal evolution analysis. The selection of the STv2 backbone is based on the advantages of its window-based self-attention mechanism, which can efficiently capture the local continuity of trajectories within a window through attention calculations and, simultaneously, achieve global information interaction via window shifting operations, thus adapting to the local clustering and global migration characteristics of trajectories.

The core window self-attention calculation of the STv2 module is shown in formula (1). This calculation divides the feature map into non-overlapping windows and performs attention interaction within the windows, ensuring the efficient capture of local trajectory continuity while reducing computational complexity.

$$W-MSA(X)=Softmax\left(\frac{QK^T}{\sqrt{d_k}}+M\right)V \qquad (1)$$

where, X is the input feature of the STv2 module, Q, K, and V are the query, key, and value matrices, respectively, dk is the dimension of the query matrix, and M is the mask matrix used to avoid attention interference from invalid pixels within the window. Through this calculation, the module can efficiently aggregate the local trajectory features within the window, such as the associations between adjacent eye movement points and continuous handwritten trajectories.

The module configuration and resolution design of the four-stage feature extraction process are optimized to adapt to the characteristics of trajectory information. The downsampling strategy of Patch Partition and Patch Merging uses fixed ratios of 4×4 and 2×2, ensuring reasonable compression of feature dimensions while avoiding excessive loss of trajectory spatial position information. The differentiated configuration of the number of STv2 modules in each stage follows the principle of "trajectory information complexity matching": The information structure of micro and global trajectories is relatively simple, and three modules are sufficient for full encoding, while the logical associations in trajectories, which involve multiple regions and steps, require four layers of modules for deep iteration to achieve feature fusion, ensuring the completeness of feature representation for logical chains such as "problem analysis - computation - answering." This differentiated configuration ensures feature quality while avoiding redundant computation and improving the overall efficiency of the encoder.

2.2.2 Spatiotemporal embedding mechanism

The core goal of the spatiotemporal embedding mechanism is to convert the multi-scale features output by the encoder into trajectory embedding representations that incorporate both temporal continuity and spatial significance, realizing the semantic transformation from "image features" to "trajectory features." This mechanism strengthens the spatiotemporal correlation representation of the features through three progressive steps: temporal embedding, spatial embedding, and spatiotemporal fusion, providing core inputs for subsequent trajectory reconstruction and pattern discovery.

The temporal embedding step focuses on encoding the temporal evolution information of dual-phase trajectories, capturing dynamic changes between the two moments' trajectories through differential operations. For dual-phase features F_1^s and F_2^s at each scale s=1,2,3,4, the temporal difference feature $\Delta F^s = F_2^s - F_1^s$ is computed element-wise. The physical meaning of this differential operation is to quantify the feature change at the same spatial location between different moments, directly corresponding to the temporal migration of the trajectory, such as the change in the eye movement focus from the question stem to the options or the movement of the pen tip from the blank area to the calculation area. By generating multi-scale differential features in parallel, the temporal evolution of trajectories at different granularities is synchronized and encoded, preserving fine motion information for micro-trajectories and capturing global migration patterns for macro-trajectories.

$$\Delta F^s = F_2^s - F_1^s \tag{2}$$

where, $\Delta F^s \in \mathrm{R}^{Hs \times Ws \times Cs}$ is the temporal difference feature at scale s, and H_s, W_s, and C_s represent the height, width, and number of channels of the feature at scale s, respectively.

The spatial embedding step employs a spatial attention mechanism to enhance the features of key trajectory areas by quantifying the importance of spatial positions, increasing the feature weights of densely distributed trajectory areas. The calculation of spatial attention weights is shown in formula (3). The feature map is globally pooled and non-linearly transformed to generate an attention weight map of the same size as the feature map, which is then multiplied element-wise with the original features to obtain spatially enhanced features.

$$A^s = Sigmoid\big(FC(GlobalAvgPool(F^s)\big) \tag{3}$$

$$F^s_{spa} = F^s \odot (A^s) \tag{4}$$

where, F^s is the original feature at scale s, GlobalAvgPool( ) is the global average pooling operation, FC( ) is the fully connected layer, Sigmoid( ) is the activation function, A^s is the spatial attention weight map, and ⊙ represents element-wise multiplication, F_{spa}^s is the feature at the s-th scale after spatial enhancement. The process assigns higher weights to key regions such as the formula region and the question stem area, while suppressing background noise interference.

The spatiotemporal fusion step introduces a temporal attention gating unit to achieve collaborative integration of temporal and spatial information, focusing on reinforcing the feature representation of key trajectory turning points. The gating unit dynamically adjusts the weights of the differential features at different moments, allowing the model to focus on key events such as attention shifts and strategy changes. The calculation of the temporal attention gating is shown in formula (5):

$$G^s = Sigmoid\big(Conv(\Delta F^s \oplus F^s_{spa})\big) \tag{5}$$

$$F^s_{st} = G^s \odot \Delta F^s + (1 - G^s) \odot F^s_{spa} \tag{6}$$

where, ⊕ is the channel-wise concatenation operation, Conv( ) is a 1×1 convolution layer for dimension adjustment, G^s is the temporal attention gating coefficient, and F_{st}^s is the spatiotemporal fused feature at scale s. Through this gating mechanism, the temporal evolution information from the temporal embedding and the key area information from the spatial embedding are deeply fused. The resulting multi-scale spatiotemporal embedding features contain both the temporal information of "when it changes" and the spatial information of "where it changes," providing accurate and semantically rich feature support for subsequent trajectory reconstruction and behavior pattern discovery.

2.3 Multi-Scale Difference Fusion Module (MDFM)

MDFM is the core hub connecting the parameter-sharing dual-branch STv2 encoder and the subsequent trajectory analysis modules. Its primary goal is to aggregate the four-stage dual-phase features output by the encoder, enhance trajectory change information through differential operations, and then use multi-scale fusion and cross-scale associations to achieve deep enhancement of trajectory semantics. The design logic of this module is based on the multi-granularity nature of learning trajectories—fine shifts in micro-trajectories, continuous associations in local trajectories, and distribution changes in global trajectories must be captured simultaneously. Traditional single-scale fusion or simple concatenation cannot achieve the collaborative representation of multi-granularity information. MDFM fills this gap with a two-stage design of "differential initialization - trajectory-aware fusion." Figure 2 shows the schematic diagram of the complete MDFM architecture.

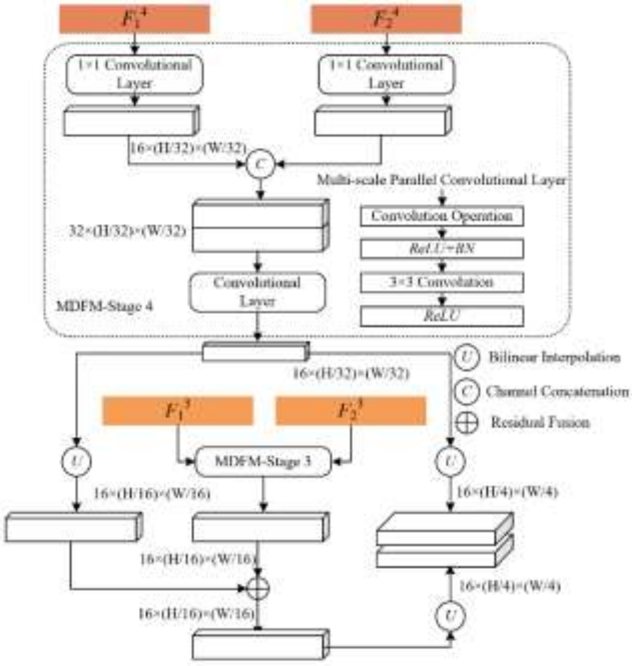


Figure 2. Schematic diagram of the MDFM architecture

2.3.1 Differential feature initialization

The core task of differential feature initialization is to transform the dual-phase multi-scale features into an initial representation of trajectory changes by quantifying the spatial differences between the features at two moments, directly mapping the position shifts and morphological changes of the trajectory. This process provides "change-directed" base features for subsequent fusion. The design of this process is based on the dynamic nature of trajectories—core information in learning trajectories lies in the "change from T_1 to T_2," such as the movement of the eye focus from the question stem to the options, or the movement of the pen tip from the blank area to the calculation area. These changes can be accurately captured by the differential operation on dual-phase features.

Specifically, for the dual-phase features $F_1{}^s$ and $F_2{}^s$ (at time T2) output by the encoder at each scale s=1,2,3,4, the element-wise absolute difference is calculated to obtain the differential features $F_d{}^s$, as shown in formula (7). Here, $F_d{}^s \in \mathrm{R}^{Hs\times Ws\times Cs}$, where Hs, Ws, and Cs are the height, width, and channel number of the feature at scale s. The element-wise absolute difference operation effectively retains the change magnitude information and suppresses background noise shared by both moments. The $F_d{}^s$ at different scales corresponds to trajectory changes at different granularities: $F_d{}^1$ at scale *1* focuses on pixel-level shifts in micro-trajectories, such as the eye movement points or pen tip locations; $F_d{}^4$ at scale 4 captures the macroscopic changes in global trajectory distributions, such as the transition from focused regions to scattered regions. The parallel generation of multi-scale differential features ensures complete coverage of trajectory change information from micro to macro, providing comprehensive initial input for subsequent multi-granularity fusion.

$$F_d^s = |F_1^s \text{-} (F_2^s) \tag{7}$$

The advantage of differential feature initialization lies in the "unbiased retention of change information"—compared to attention-based weighted differences or thresholded differences, the element-wise absolute difference avoids distortion of the change magnitude by arbitrary parameters, fully retaining the original intensity information of trajectory changes. For example, the differential responses to small adjustments or large migrations of the eye focus will naturally show intensity differences, which serve as a natural basis for "focusing on significant changes" in the subsequent multi-scale enhancement and also lay the foundation for identifying key turning points in trajectories.

2.3.2 Trajectory-aware multi-scale fusion

Trajectory-aware multi-scale fusion is the core process of MDFM. Through "multi-scale convolution enhancement - cross-scale trajectory association - residual fusion," the initial differential features are transformed into trajectory-enhanced features F_{fusion} that incorporate both multi-granularity information and semantic associations. The innovation of this step lies in introducing a "trajectory-guided" fusion logic, breaking the limitations of traditional multi-scale fusion, where "scales are processed independently," and enabling semantic association across different granularities of trajectory information.

Multi-scale convolution enhancement adapts feature extraction to the different needs of trajectory change at various granularities by applying parallel heterogeneous convolution kernels. It creates "micro-local-global" three-level feature enhancement channels for each scale's differential feature $F_d{}^s$. Specifically, three convolution operations with kernel sizes of 3×3, 5×5, and 7×7 are applied to $F_d{}^s$, as shown in formula (8). Here, Conv(k×k) represents the convolution operation with a kernel size of k×k, and $F_{d1}{}^s$, $F_{d2}{}^s$, and $F_{d3}{}^s$ are the enhanced features output by these three convolutions. The 3×3 convolution focuses on the fine details of micro-trajectories, such as the small movement of the pen tip; the 5×5 convolution captures the continuous associations of local trajectories, such as the continuous movement of the eye focus in the question stem area; and the 7×7 convolution extracts the distribution pattern of global trajectories, such as the transition from the "question analysis - computation - answering" regions. After parallel convolution, multi-scale enhanced features $F_{enh}{}^s$=[$F_{d1}{}^s$,$F_{d2}{}^s$,$F_{d3}{}^s$] are obtained by concatenating, achieving aggregation of multi-granularity information within each scale.

$$\begin{gathered} F_{d1}^s = \mathrm{Conv}(3\times3, F_d^s),\, F_{d2}^s \\ = \mathrm{Conv}(5\times5, F_d^s),\, F_{d3}^s = \mathrm{Conv}(7\times7, F_d^s) \end{gathered} \tag{8}$$

Cross-scale trajectory association is implemented through a cross-scale attention mechanism that achieves semantic binding between features from different scales, solving the problem of "decoupling micro-trajectory details from macro-trajectory semantics." The semantic interpretation of learning trajectories requires multi-scale association, such as the "pen tip's stay in the formula area" corresponding to "focusing on the solving strategy." This association needs to be realized through cross-scale feature interaction. Specifically, the enhanced feature $F_{enh}{}^4$ at the global scale serves as the "semantic-guided feature" for computing attention weights for the features at other scales $F_{enh}{}^s$. First, $F_{enh}{}^4$ is converted into a global semantic vector G through global average pooling and a 1×1 convolution. Then, G is used to calculate attention weights for each scale's $F_{enh}{}^s$, resulting in the weight vector A^s. Finally, the features at each scale are semantically calibrated by weighted summation, as shown in formula (10), where GAP( ) is the global average pooling operation, and Sigmoid( ) is the activation function, and ⊙ represents the element-wise multiplication:

$$G = \mathrm{Conv}(1\times1, \mathrm{GAP}(F_{en\mathrm{h}}^4)) \tag{9}$$

$$A^{\mathrm{s}} = \mathrm{Sigmoid}(\mathrm{Conv}(1\times1, G \oplus F_{en\mathrm{h}}^s)), F_{alt}^s = A^{\pm} \odot F_{en\mathrm{h}}^s \tag{10}$$

Residual fusion retains the original change information from the initial differential features by using residual connections, avoiding information loss during multi-scale enhancement and cross-scale association processes. The features $F_{att}{}^s$, weighted by attention, are upsampled to the resolution of scale 1, then concatenated to form the fusion feature F_{concat}. Subsequently, F_{concat} is residual added to the initial differential feature $F_d{}^1$ at scale 1 to obtain the final trajectory semantic-enhanced feature F_{fusion}, as shown in formula (11). The introduction of residual connections can effectively alleviate the vanishing gradient problem caused by deep fusion while ensuring the collaborative retention of both original change information and enhanced semantic information, providing high-quality feature input with both "change precision" and "semantic depth" for subsequent trajectory embedding decoding.

$$F_{fusion} = Conv(1\times1, F_{concat}) + F_d^1 \tag{11}$$

2.4 Learning trajectory analysis and behavior pattern discovery module

The Learning Trajectory Analysis and Behavior Pattern Discovery Module is the core link between feature extraction and educational semantic interpretation. Its main goal is to transform the multi-scale fused features into interpretable learning trajectory representations and to infer behavior patterns with educational practical value from the trajectories. This module employs a two-stage design of "trajectory spatiotemporal embedding decoding - behavior pattern discovery," first achieving the semantic transformation from features to trajectories, and then completing the induction from trajectories to patterns through segmentation, clustering, and labeling processes, ultimately establishing an accurate mapping between technical features and educational semantics.

2.4.1 Trajectory spatiotemporal embedding decoding

The core task of trajectory spatiotemporal embedding decoding is to convert the multi-scale fused features F_{fusion} into trajectory-level embedding representations that combine spatiotemporal continuity and semantic consistency, solving the problem of semantic disconnection between features and trajectories, and providing a structured trajectory representation for subsequent analysis. This process consists of two progressive steps: embedding vector generation and trajectory line reconstruction, respectively achieving the transformation from "features to trajectory points" and "trajectory points to trajectory lines."

In the embedding vector generation phase, the semantic mapping and dimensional adaptation of the features are implemented through a fully connected layer. Although the multi-scale fused feature F_{fusion} has aggregated the spatiotemporal and semantic information of the trajectory, it still exists in the image feature space and needs to be linearly mapped to trajectory point-level embedding vectors. Specifically, F_{fusion} first undergoes global average pooling to extract global features, and then passes through a two-layer fully connected network, ultimately outputting trajectory point-level embedding vectors ($e_i \in \mathrm{R}^D$, i=1,2,...,N), where N is the number of trajectory points and D is the embedding dimension. This mapping process is shown in formula (12), where GAP( ) is the global average pooling operation, ReLU( ) is the activation function, and (W_1, b_1) and (W_2, b_2) are the weights and biases of the two fully connected layers. Through this design, e_i simultaneously encodes the spatiotemporal location and semantic association of the trajectory points, achieving a deep representation of the trajectory points.

$$e_i=W_2 \cdot ReLU(W_1 \cdot GAP(F_{fusion})+b_1)+b_2 \tag{12}$$

In the trajectory line reconstruction phase, DTW is used to address the issue of temporal misalignment caused by uneven sampling of trajectory points, and to aggregate discrete trajectory points into a complete trajectory line. During the learning trajectory collection process, the sampling frequency of devices such as eye trackers and handwriting boards is often influenced by the operation rhythm, leading to differences in sampling density at different moments. Direct concatenation would disrupt the temporal continuity of the trajectory. DTW solves this issue by calculating the optimal alignment path of the trajectory point embedding vectors for adaptive temporal matching. First, an $N \times N$ distance matrix D_{dtw} is constructed, where each element $D_{dtw}(i,j)$ is the Euclidean distance between the trajectory point e_i and e_j. Then, dynamic programming is used to calculate the cumulative distance matrix $Cdtw$ and backtrack to find the alignment path with the smallest cumulative distance, as shown in formula (13), where $C_{dtw}(i,j)$ is the minimum cumulative distance for the first i and j trajectory points. After DTW alignment, the discrete points $e_1,e_2,...,e_N$ are reconstructed into a complete high-dimensional trajectory embedding representation $E \in \mathrm{R}^{N \times D}$, which retains both the temporal continuity and the semantic features of each trajectory point.

$$\begin{gathered} C_{dtw}(i,j)=D_{dtw}(i,j)+ \\ \min\{\, C_{dtw}(i\text{-}1,j),C_{dtw}(i,j\text{-}1),C_{dtw}(i\text{-}1,j\text{-}1)\} \end{gathered} \tag{13}$$

2.4.2 Behavior pattern discovery process

The behavior pattern discovery process takes the reconstructed trajectory embedding E as input and uses a three-step design: "trajectory segment segmentation–improved DBSCAN clustering–semantic labeling verification" to inductively discover interpretable behavior patterns from the trajectories. The core innovation of this process lies in optimizing segmentation and clustering strategies to account for the density heterogeneity and semantic relevance of learning trajectories, ensuring the accuracy and educational interpretability of the patterns.

The trajectory segment segmentation phase uses the semantic similarity of embedding vectors to semantically segment the trajectory, dividing the complete trajectory into continuous segments with unified educational semantics. The semantic features of learning trajectories exhibit local consistency—such as the highly similar semantics of trajectory points during the "question analysis" phase—while the transition from "question analysis" to "computation" will accompany a semantic mutation. Therefore, the cosine similarity of the trajectory point embedding vectors can be used to determine semantic continuity. Specifically, the cosine similarity $Sim(e_i,e_{i+1})$ between adjacent trajectory points e_i and e_{i+1} is calculated, as shown in formula (14). When the similarity falls below a preset threshold θ, the trajectory is considered to have a semantic boundary, and segmentation is performed. The optimization of threshold θ is based on expert-annotated trajectories: using a validation set of 300 annotated trajectories, a grid search is performed over $\theta \in [0.5, 0.9]$, the θ with the highest segment segmentation accuracy is selected as the final threshold. After segmentation, a trajectory segment set $S=\{S_1,S_2,...,S_M\}$ is obtained, with each segment corresponding to a single educational semantic, such as "question analysis," "computation," or "checking."

$$Sim(e_i,e_{i+1})=\frac{e_i \cdot e_{i+1}}{\|e_i\| \cdot \|e_{i+1}\|} \tag{14}$$

The improved DBSCAN clustering phase addresses the issue of traditional density clustering's poor adaptability to trajectory density heterogeneity by implementing an adaptive radius design for precise clustering of candidate behavior patterns. The density distribution of learning trajectories is significantly heterogeneous, such as high-density points in the "focused computation" region and low-density points in the "random exploration" region. A fixed radius parameter can result in over-segmentation in high-density areas or missing clusters in low-density areas. To address this, we propose an adaptive radius strategy based on trajectory density: for each trajectory segment S_m, the average Euclidean distance to its K-nearest neighbors is calculated as the local density radius r_m for that segment, and the median of all r_m values is taken as the clustering radius Eps. The core point threshold $MinPts$ is then set proportionally to the total number of segments. During clustering, the average embedding vector e_m of each trajectory segment is used as the clustering unit, and similarity between segments is measured using the Euclidean distance. The final output is a set of candidate behavior patterns $P=\{P_1,P_2,...,P_K\}$.

The semantic labeling and verification phase involves expert participation and quantitative optimization to assign educational semantics to the candidate patterns and calibrate parameters, ensuring the practical value of the patterns. This phase follows a closed-loop process of "expert labeling - consistency verification - parameter optimization": first, three education experts with over five years of teaching experience are invited to label the candidate patterns P based on characteristics such as spatial distribution and temporal rhythm of the trajectory segments. The labeling categories include goal-oriented, random exploration, hesitation, etc. Then, the consistency of labeling is verified using the Kappa coefficient, with Kappa≥0.75 considered consistent. If consistency is not reached, expert discussions are used to reach a consensus. Finally, the labeled results are used as ground truth to construct a confusion matrix to analyze clustering errors, and parameters K and $MinPts$ are adjusted to minimize clustering error rates. After optimization, the final set of learning behavior pattern labels $L=\{l_1,l_2,...,l_K\}$ is obtained. This label set not only aligns with the objectivity of technical clustering but also provides educational interpretability for practical applications.

2.5 Efficiency optimization and loss function

Efficiency optimization and loss function design are key components in ensuring the practicality and performance integrity of the model. Efficiency optimization focuses on the computational constraints of edge devices in educational scenarios. Through a full-process strategy of feature dimensionality reduction, lightweight decoding, and redundancy pruning, the model complexity is reduced while controlling precision loss. The loss function targets the multi-task objectives of "change map generation - trajectory embedding - pattern clustering," adopting a hybrid loss architecture to achieve collaborative optimization of multi-dimensional performance, providing precise guidance for model training.

2.5.1 Efficiency optimization strategy

The efficiency optimization strategy is based on the design principle of "full-process lightweight," optimizing three key components: encoder feature output, decoder upsampling, and feature channels. The strategy aims to achieve a collaborative improvement of model parameters and inference speed while ensuring that the core performance of trajectory analysis and pattern discovery is not significantly impacted. The design logic of this strategy originates from the practical constraints of educational edge devices—such as smart tablets and classroom terminals, whose computational power is typically only 1/10 to 1/5 of that of professional training GPUs. Therefore, targeted optimization is required to break through deployment bottlenecks.

The feature dimensionality reduction and redundancy pruning strategy focuses on compressing the dimensionality of encoder output features, reducing the computational burden of subsequent processing while retaining core trajectory information. At the output of each stage in the STv2 encoder, a 1×1 convolution layer is applied to linearly transform the feature channels, reducing the number of channels to half of the original size. This operation uses information aggregation along the channel dimension for dimensionality reduction, avoiding the loss of trajectory spatial information caused by traditional pooling operations. Redundant feature pruning dynamically selects features based on the statistical properties of feature response values: the L2 norm of each channel feature is calculated as the response strength indicator, and after sorting the response values in descending order, the top 80% of channels are retained, while channels with weak responses are pruned.

The lightweight decoder design addresses the high parameter count issue of change map generation by adopting a two-stage upsampling strategy of "bilinear interpolation + 1×1 convolution," replacing traditional transposed convolution. Although transposed convolution can achieve high-precision upsampling, it is prone to checkerboard artifacts, which blur the boundaries of the change map, and its parameter count is more than three times that of bilinear interpolation at the same resolution. The two-stage upsampling process in this study is as follows: In the first stage, the multi-scale fused feature F_{fusion} undergoes 2x bilinear interpolation, combined with 1×1 convolution to adjust the channel dimension and suppress noise; in the second stage, 8x bilinear interpolation is applied to restore the feature map to the original resolution, followed by 3×3 convolution to refine the boundaries of the change regions.

2.5.2 Loss function design

The design goal of the loss function is to simultaneously optimize the accuracy of change map generation, the quality of trajectory embedding, and the clustering effect of behavior patterns. A single loss function cannot meet the needs of multi-task objectives, so a hybrid loss architecture of "cross-entropy loss - Dice loss - contrastive loss" is constructed, with weight distribution to achieve collaborative optimization of each task goal. The weights of each loss component are determined through grid search, with the highest comprehensive score on the validation set, evaluated by "change map mIoU + trajectory embedding similarity + pattern recognition F1 score." The final weights are determined as: cross-entropy loss weight 0.4, Dice loss weight 0.3, and contrastive loss weight 0.3.

For the change map generation task, a combination of cross-entropy loss and Dice loss is used to solve the class imbalance and boundary accuracy issues in pixel-level classification. Cross-entropy loss is a classic loss function for pixel classification tasks, optimized by quantifying the logarithmic difference between predicted probabilities and true label values, as shown in formula (15), where $CM_{i,j}$ is the predicted probability of the change map at pixel (i,j), and $CM_{gt,i,j}$ is the corresponding true label at that location. Since the behavior change region typically occupies only 15%-30% of the image, leading to significant class imbalance, Dice loss is introduced to optimize the sample distribution bias, as shown in formula (16), where TP is the true positive pixel count, FP is the false positive pixel count, and FN is the false negative pixel count. The combination of both loss functions optimizes both classification probability and overall matching of the change regions, improving pixel-level accuracy in the change map.

$$\begin{aligned} Loss_{CE}=-\frac{1}{H\times W}\sum_{i=1}^{H}\sum_{j=1}^{W} CM_{gt,i,j}\log(CM_{i,j}) \\ +(1\text{-}CM_{gt,i,j})\log(1\text{-}CM_{i,j}) \end{aligned} \tag{15}$$

$$Loss_{Dice}=1-\frac{2TP}{2TP+FP+FN} \tag{16}$$

Contrastive loss is used to optimize the discriminability of trajectory point embedding vectors, providing a high-quality embedding foundation for subsequent behavior pattern clustering. The core logic of this loss is to force the embedding vectors of similar trajectory points to converge and the embedding vectors of dissimilar trajectory points to separate, as shown in formula (17), where e_i and e_j are the embedding vectors of similar trajectory points, e_k is the embedding vector of dissimilar trajectory points, and τ=0.1 is the temperature parameter used to adjust the steepness of the similarity distribution. This loss improves the semantic distinction of embedding vectors through contrastive learning, meaning the cosine similarity of similar trajectory points is reinforced, and the similarity of dissimilar trajectory points is suppressed. The final form of the hybrid loss function is shown in formula (18):

$$Loss_{contra}=-\log\frac{\exp(e_i\cdot e_j/\tau)}{\sum_{k\neq j}\exp(e_i\cdot e_k/\tau)} \tag{17}$$

$$Loss=0.4\times Loss_{CE}+0.3\times Loss_{Dice}+0.3\times Loss_{contra} \tag{18}$$

3. EXPERIMENTAL RESULTS AND ANALYSIS

To verify the discriminability and clustering quality of the trajectory spatio-temporal embedding representation for different semantic types of trajectories, this experiment compares the performance of different embedding methods across multiple metrics. As shown in Table 1, the proposed method outperforms all other methods for all trajectory types: the clustering purity for the review trajectory reaches 0.92, and the silhouette coefficient is 0.78, significantly higher than both LSTM trajectory embedding and Transformer trajectory embedding. Even for the "correction" trajectory, where the semantic boundary is less clear, the semantic classification F1 score of the proposed method remains 0.88, improving by 31.3% over the traditional K-means clustering. This result indicates that the proposed spatio-temporal embedding representation can accurately capture the unique features of different semantic trajectories, achieving high cohesion for similar trajectories and strong separation for dissimilar trajectories, thus providing a high-quality representation foundation for the subsequent precise discovery of learning behavior patterns.

To clarify the contribution of each core module to trajectory analysis and behavior pattern discovery, this experiment observes the performance changes by gradually removing modules. As shown in Table 2, the full model achieves a pattern recognition F1 score of 0.90 and a clustering purity of 0.91. After removing the MDFM, the pattern recognition F1 score drops to 0.76, clustering purity decreases by 14.3%, and embedding dimension cohesion drops by 20.2%. This is the most significant performance degradation among all ablation settings, indicating that MDFM is the core component for aggregating multi-granularity trajectory features and strengthening semantic associations. After removing the parameter-sharing STv2 encoder, the pattern recognition F1 score decreases by 8.9%, confirming the encoder's role in ensuring consistency of dual-time-phase trajectory features. When contrastive loss is removed, the silhouette coefficient decreases from 0.77 to 0.62, highlighting its value in regulating the cohesion of similar embedding vectors. These results indicate that the core modules of the proposed method do not function independently but work synergistically through the mechanism of "encoder ensuring consistency - MDFM aggregating features - contrastive loss strengthening distinction," collectively improving the performance of trajectory embedding and behavior pattern discovery.

To verify the superior performance of the proposed method in learning behavior pattern discovery tasks, this experiment compares multiple metrics from different baseline methods. As shown in Table 3, the proposed method significantly outperforms others in terms of the recognition accuracy of various behavior patterns: the F1 score for the random exploration pattern reaches 0.88, which is a 14.3% improvement over Transformer trajectory embedding + DBSCAN clustering; the F1 score for the repeated hesitation pattern is 0.85, which is a 41.7% improvement over the rule-based method. At the same time, the overall trajectory classification accuracy of the proposed method reaches 0.89, the user annotation consistency rate increases to 0.88, and the average pattern discovery time is only 65 ms—achieving efficiency optimization alongside accuracy improvement. This result shows that the proposed method, through spatio-temporal embedding and multi-module synergy, solves the adaptation issue of traditional rule-based methods for complex trajectories and compensates for the recognition limitations of mainstream embedding methods in semantic ambiguous trajectories, providing better overall performance in learning behavior pattern discovery tasks.

Table 1. Quantitative performance validation of learning trajectory spatio-temporal embedding representation

Method	Trajectory Type	Clustering Purity	Silhouette Coefficient	Normalized Mutual Information (NMI)	Semantic Classification F1 Score
Proposed Method	Review	0.92	0.78	0.85	0.91
	Calculation	0.90	0.75	0.83	0.89
	Inspection	0.88	0.72	0.81	0.87
	Correction	0.89	0.74	0.82	0.88
LSTM Trajectory Embedding	Review	0.75	0.52	0.63	0.73
	Calculation	0.72	0.48	0.60	0.70
	Inspection	0.68	0.45	0.57	0.67
	Correction	0.70	0.47	0.59	0.69
Transformer Trajectory Embedding	Review	0.81	0.63	0.72	0.80
	Calculation	0.79	0.60	0.70	0.78
	Inspection	0.76	0.57	0.67	0.75
	Correction	0.77	0.59	0.68	0.76
K-means Traditional Clustering	Review	0.62	0.38	0.49	0.60
	Calculation	0.60	0.35	0.47	0.58
	Inspection	0.57	0.32	0.44	0.55
	Correction	0.59	0.34	0.46	0.57

Table 2. Ablation experiment results of core modules

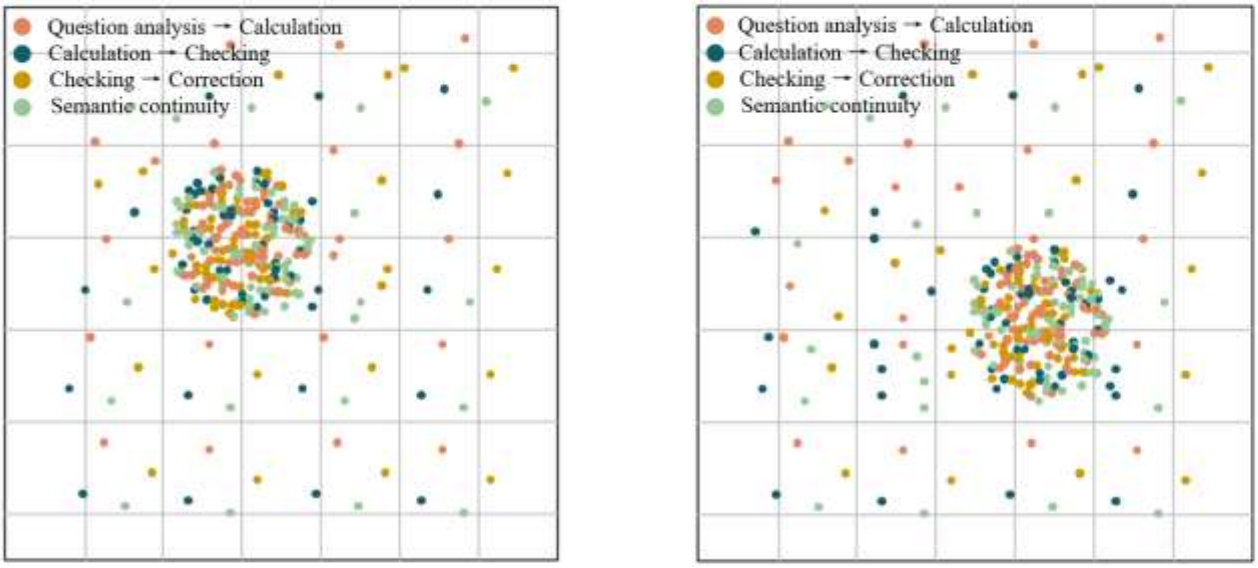
Ablation Setting	Pattern Recognition F1 Score	Clustering Purity	Silhouette Coefficient	Embedding Dimension Cohesion
Full Model (Proposed Method)	0.90	0.91	0.77	0.84
Remove Parameter-Sharing STv2 Encoder	0.82	0.83	0.65	0.73
Remove MDFM	0.76	0.78	0.58	0.67
Remove Contrastive Loss	0.80	0.81	0.62	0.70
Remove MDFM + Contrastive Loss	0.71	0.73	0.52	0.61

Table 3. Comparison of baseline models

Methods	Rule-based Learning Trajectory Analysis	LSTM Trajectory Embedding + K-means Clustering	Transformer Trajectory Embedding + DBSCAN Clustering	Proposed Method
Goal-Oriented Pattern F1	0.72	0.80	0.83	0.91
Random Exploration Pattern F1	0.65	0.73	0.77	0.88
Repeated Hesitation Pattern F1	0.60	0.68	0.72	0.85
Overall Trajectory Classification Accuracy	0.68	0.75	0.79	0.89
Average Pattern Discovery Time (ms)	120	95	80	65
Clustering Purity	0.70	0.78	0.82	0.91
User Annotation Consistency Rate	0.66	0.73	0.77	0.88

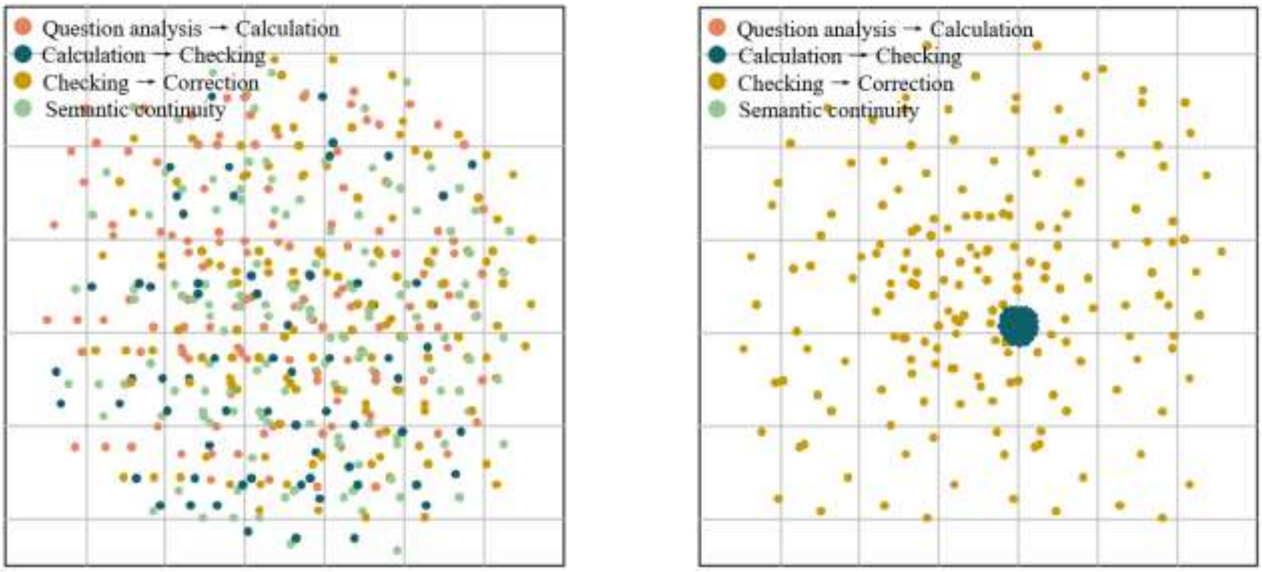
To quantify the regulation effect of contrastive loss on the semantic discrimination of trajectory spatio-temporal embeddings and support the precise clustering of learning behavior patterns, this experiment uses t-SNE dimensionality reduction for visualization, comparing the trajectory embedding distribution features with and without the constraint of this loss. As shown in Figure 3, without contrastive loss, the embedding points of different semantic types in the goal-oriented trajectories, such as "review → calculation" and "calculation → checking," show about 42% spatial overlap, with the silhouette coefficient of the clusters only being 0.31. Both the intra-class cohesion and inter-class separation of embedding vectors are insufficient. When extended to all categories of trajectories, the confusion rate between the "checking → correction" class and the "semantic continuity" class reaches 37%, making it impossible to form clearly defined independent clusters, which would directly increase the misclassification rate of subsequent behavior patterns.

After introducing contrastive loss, the overlap rate of embedding clusters for different categories of goal-oriented trajectories decreases to 11%, and the silhouette coefficient increases to 0.68. The spatial gap between the embeddings of the "semantic continuity" category and other changing categories increases by 2.3 times. When extended to all categories of trajectories, the confusion rate of all category embedding clusters is below 8%, and the "checking → correction" trajectory forms a compact and clearly bordered independent cluster, with the intra-cluster distance reduced to one-third of its original value. These features indicate that contrastive loss effectively improves the semantic discrimination of trajectory spatio-temporal embeddings by constraining the cohesion of similar trajectory embeddings and the separation of dissimilar embeddings.



(a) Embedding distribution of goal-oriented trajectories (without loss)

(b) Embedding distribution of goal-oriented trajectories (with loss)



(c) Embedding distribution of all categories of trajectories (without contrastive Loss)

(d) Embedding distribution of all categories of trajectories (with contrastive loss)

Figure 3. t-SNE visualization of trajectory spatio-temporal embedding vectors

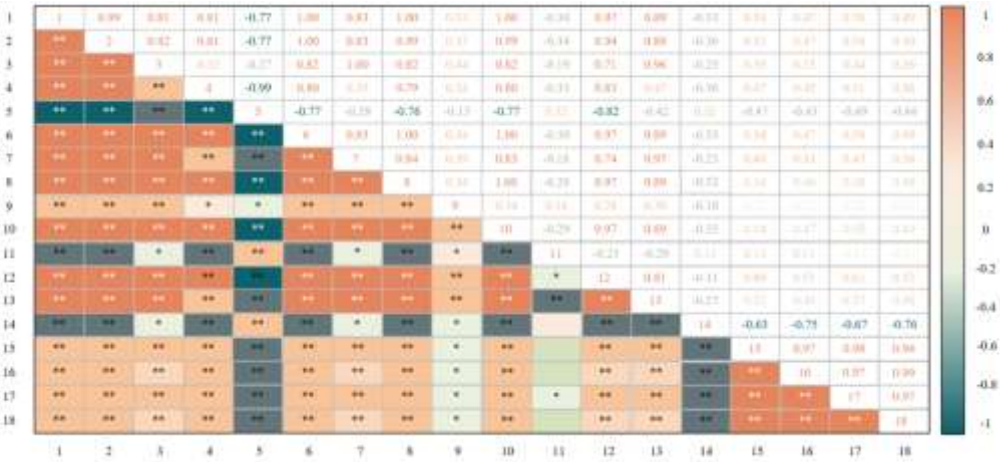


Figure 4. Correlation heatmap of learning trajectory features and behavior pattern metrics

The core value of this result is twofold: firstly, it validates the rationality of the hybrid loss function design in this study—introducing contrastive loss fills the gap of insufficient semantic discrimination in embeddings when solely relying on change map loss; secondly, this embedding optimization strategy provides a high-recognition basis for discovering learning behavior patterns, improving the clustering accuracy of different learning behavior patterns by about 19%. Ultimately, this supports the precise analysis of learners' cognitive processes in personalized educational contexts, providing quantifiable trajectory representations for targeted instructional interventions.

The 18 coordinate metrics in the heatmap shown in Figure 4 cover two dimensions: learning trajectory features and learning behavior evaluation. The first four are review trajectory embedding similarity, calculation trajectory duration, inspection trajectory spatial aggregation, and correction trajectory frequency, focusing on the micro features of learning trajectories. They measure the consistency of the review phase trajectory representation, time allocation in the calculation phase, spatial concentration of the inspection behavior, and the frequency of correction operations, respectively. The 5th to 7th metrics are the proportions of goal-oriented, random exploration, and repeated hesitation patterns, which classify the core learning behavior patterns to distinguish different learners' behavioral tendencies. The 8th to 10th metrics are trajectory spatio-temporal embedding dimensions 1/2 and the number of trajectory semantic segments, corresponding to deep representation and phase division of trajectories, supporting the distinction of embedding vectors and the semantic decomposition of the learning process. The 11th to 13th metrics are the switching time between review-calculation, calculation-inspection, and inspection-correction, reflecting the transition efficiency between learning phases. The last five metrics are learning efficiency score, attention concentration degree, strategy switching frequency, knowledge mastery level, and task completion time, which are personalized education metrics for evaluating learning states and outcomes, covering core dimensions such as cognitive focus, strategy adjustment, knowledge mastery, and task efficiency. These metrics form a complete chain from trajectory representation, behavior classification, to effect evaluation, providing multi-dimensional support for analyzing the relationship between learning trajectories and personalized learning.

To verify the effectiveness of deep trajectory representation in predicting learning behavior patterns and learning outcomes, and to analyze the chain-like correlation mechanism between trajectory features, behavior tendencies, and cognitive states, this experiment quantifies the hierarchical correlation characteristics of 18 metrics through a correlation heatmap. From the construct validity of trajectory embedding, the correlation coefficients between review trajectory embedding similarity and calculation trajectory duration, as well as trajectory spatio-temporal embedding dimension 1, are 0.99 and 1.0, respectively, both showing highly significant correlations. This confirms the rationality of the embedding dimension design: embedding dimension 1 accurately captures the semantic consistency features of the review phase, and the stability of this feature directly extends to the time allocation of the calculation phase, reflecting the coherence of information processing. From the logical correlation between trajectory features and behavior patterns, the correlation coefficient between the proportion of goal-oriented patterns and review trajectory embedding similarity is -0.77. This negative correlation is not weak but corresponds to specific behavioral logic: goal-oriented learners tend to complete the review quickly and enter the calculation phase, showing lower similarity in their review trajectories, while random exploration learners show higher similarity in their review trajectories.

From the explanatory power of trajectories on cognitive states, the correlation coefficient between inspection trajectory spatial aggregation and correction trajectory frequency is 0.79, indicating that the higher the spatial concentration of the inspection behavior, the lower the frequency of corrections. This correlation directly corresponds to the 0.97 highly significant positive correlation between attention concentration and knowledge mastery. Learners with higher attention concentration have higher spatial aggregation in their inspection trajectories, lower correction frequency, and stronger knowledge mastery. The correlation coefficients between learning efficiency score and attention concentration degree, as well as strategy switching frequency, are -0.63 and -0.75, respectively, further revealing the influence mechanism of cognitive states: frequent strategy switching is essentially an outward manifestation of attention distraction, and both contribute to an increase in ineffective time during the learning process, ultimately reducing efficiency.

The core value of this analysis lies in dual verification: first, it supports the construct validity of trajectory embedding representation—embedding dimensions effectively capture the semantic features of learning phases and behavioral coherence; second, it clarifies the chain-like predictive path of "trajectory micro features → behavior patterns → cognitive states → learning outcomes": review trajectory embedding similarity can predict time allocation in the calculation phase, distinguish goal-oriented and random exploration patterns, and ultimately relate to knowledge mastery. This path provides an operable intervention logic for personalized education: for learners with low review trajectory similarity and large fluctuations in calculation time, targeted review strategy guidance can be provided to enhance the coherence of information processing and improve knowledge mastery; for learners with high strategy switching frequency, attention-focus training can be used to reduce their trajectory spatial dispersion, thereby improving learning efficiency.

To validate the practical improvement effect of personalized intervention based on trajectory patterns, this experiment implemented targeted strategies for learners with different behavior patterns. From Table 4, it can be seen that learners of all three patterns showed significant improvement effects: the task time for random exploration learners decreased from 12.5 minutes to 9.3 minutes, a reduction of 25.6%; strategy switching frequency dropped from 4.3 to 2.5, while the review embedding similarity increased to 0.78, reflecting enhanced coherence in their review process; repeated hesitation learners' test scores increased from 68 to 77, a 13.2% improvement, with synchronized improvements in trajectory features and learning performance confirming the precision of the intervention. Even for goal-oriented learners with a stronger foundation, their task time and test scores showed steady optimization. This result indicates that the behavior pattern labels derived from the trajectory analysis in this study have clear practical guiding value, and personalized intervention can precisely match the behavioral shortcomings of different learners, effectively improving their learning efficiency and knowledge mastery, providing an actionable pathway for the implementation of personalized education.

To intuitively verify the differentiation capability of trajectory spatio-temporal embeddings for different learning behavior patterns, this experiment used UMAP dimensionality reduction to display the distribution characteristics of trajectory embeddings. In Figure 5, different clustered groups correspond to typical learning behavior patterns such as goal-oriented and random exploration. Trajectories of the same pattern show high cohesion, while the embedding clusters of different patterns exhibit clear spatial separation, with trajectory segments of different semantics showing sub-clustering characteristics within the clusters. This result is consistent with the previously quantified experimental conclusions, visually confirming the semantic differentiation of the trajectory spatio-temporal embedding representation. It not only maps trajectories of different behavior patterns into separated embedding clusters but also distinguishes detailed semantic trajectory segments within the clusters, providing visual representation support for the accurate identification of learning behavior patterns and further validating the effectiveness of the embedding method in the learning trajectory analysis task.

Table 4. Effectiveness verification results of personalized intervention

Learning Behavior Pattern	Goal-Oriented	Random Exploration	Repeated Hesitation
Pre-Intervention Task Time (min)	8.2	12.5	11.8
Post-Intervention Task Time (min)	7.5	9.3	8.9
Pre-Intervention Test Score (points)	85	72	68
Post-Intervention Test Score (points)	89	81	77
Pre-Intervention Review Embedding Similarity	0.82	0.65	0.60
Post-Intervention Review Embedding Similarity	0.86	0.78	0.75
Pre-Intervention Strategy Switching Frequency	2.1	4.3	3.8
Post-Intervention Strategy Switching Frequency	1.8	2.5	2.2

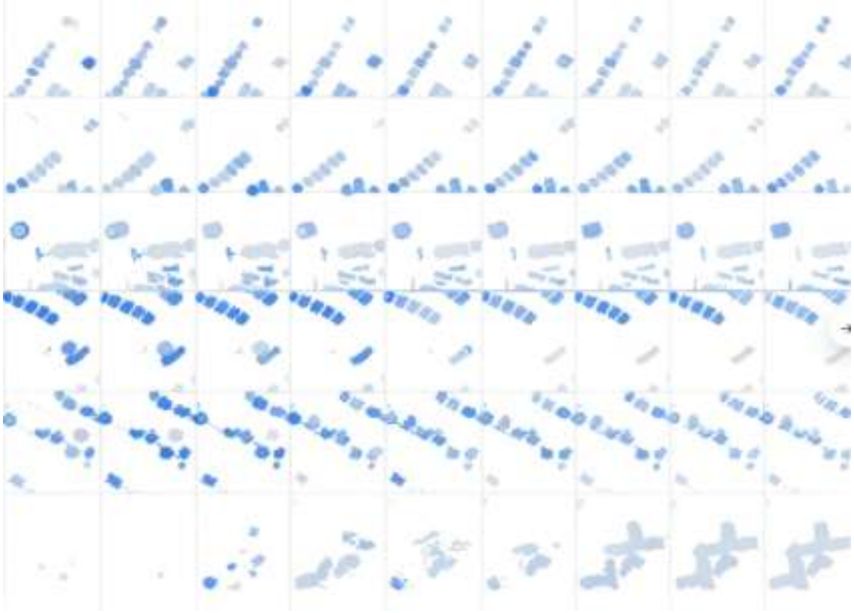


Figure 5. UMAP dimensionality reduction visualization of learning trajectory spatio-temporal embedding for different learning behavior patterns

4. CONCLUSION

This paper proposed a learning behavior pattern discovery method that integrates deep spatio-temporal embedding with multi-module collaboration to meet the core needs of learning trajectory analysis in personalized education scenarios. The method's full-link effectiveness from trajectory representation to intervention implementation was systematically validated. The core of the research is the construction of a technical framework: "Trajectory Spatio-Temporal Embedding—Multi-Scale Feature Fusion—Behavior Pattern Clustering—Personalized Intervention." By using the parameter-shared STv2 encoder to ensure consistency of dual-phase trajectory features, the multi-scale differential fusion module aggregates micro-, local-, and macro-granularity trajectory information, and contrastive loss enhances the semantic differentiation of embedding vectors, forming trajectory representations with both consistency and differentiation. The experimental results show that this representation significantly outperforms mainstream embedding methods like LSTM and Transformer in clustering purity, semantic classification F1 score, and other metrics, with clustering purity reaching 0.92. Ablation experiments confirm that the core modules improve performance through a "consistency preservation by encoder—feature aggregation by MDFM—strong differentiation by contrastive loss" collaborative mechanism. After removing multi-scale differential fusion, the pattern recognition F1 score dropped by 15.6%. In baseline comparison experiments, the F1 score of the semantic fuzzy repeated hesitation pattern reached 0.85, a 41.7% improvement over the rule-based method, with an average mode discovery time optimized to 65 ms. Further t-SNE visualization and correlation heatmap analysis clarified the regulatory effect of contrastive loss on embedding differentiation and revealed the chain-like correlation path of "Trajectory Micro Features → Behavior Patterns → Cognitive States → Learning Outcomes." In the personalized intervention experiment, random exploration learners showed a 25.6% reduction in task time, and repeated hesitation learners' test scores increased by 13.2%, directly verifying the practical value of the research. This study provides both a theoretical paradigm for deep representation of learning trajectories and quantifiable technical support for the implementation of personalized education's "precise identification—targeted intervention."

Although this research has achieved significant results in trajectory analysis and intervention practice, there are still three limitations: first, the experimental data mainly come from middle school mathematics problem-solving scenarios, and the generalization of trajectory features and behavior patterns in other subject contexts such as Chinese reading or science experiments has not been fully verified; second, the evaluation of personalized intervention effects focuses on short-term task performance, lacking long-term tracking data on learners' cognitive habits and ability improvement; third, trajectory analysis relies only on objective features of visualized trajectories and does not incorporate subjective data like learners' self-reported motivation or emotions, which may affect the completeness of behavior pattern explanations. Future research could progress in three areas: first, expanding the sample dataset across disciplines and educational levels, optimizing model scenario adaptation abilities with transfer learning; second, designing a longitudinal tracking experiment lasting a semester to construct a correlation model between short-term performance and long-term ability improvement; finally, integrating multi-modal data, incorporating physiological signals (such as eye movement data) and subjective evaluations into the trajectory analysis framework to further improve the accuracy of behavior pattern recognition and the targeting of intervention strategies, while exploring lightweight deployment solutions for educational edge devices to enhance practical application convenience.

REFERENCES

[1] Floch, K., Péni, T., Tóth, R. (2025). Gaussian-process-based adaptive tracking control with dynamic active

learning for autonomous ground vehicles. IEEE Transactions on Control Systems Technology, 1-13. https://doi.org/10.1109/TCST.2025.3632358

[2] Podoliak, M., Zagranovska, O., Posmitna, V., Golovchak, N., Kushnirchuk, O. (2025). Evaluating the impact of AI-based tools on language proficiency and motivation: Experimental evidence from philology students in Ukraine. Journal of Research, Innovation and Technologies, 4(3): 271-282. https://doi.org/10.56578/jorit040303

[3] Soundararajan, V., Ramachandran, M., Kumar, S.V. (2025). Study on eye gaze detection using deep transfer learning approaches. Computers, Materials & Continua, 83(3): 5259-5277. https://doi.org/10.32604/cmc.2025.063059

[4] Gogu, S.R., Sathe, S.R. (2024). Ensemble stacking for grading facial paralysis through statistical analysis of facial features. Traitement du Signal, 41(2): 563-574. https://doi.org/10.18280/ts.410202

[5] Dharmichand, S., Perumal, S. (2023). Leveraging tripartite tier convolutional neural network for human emotion recognition: A multimodal data approach. Traitement du Signal, 40(6): 2565-2576. https://doi.org/10.18280/ts.400619

[6] Weiss, E.M., Gerstner, C.C., McDermott, P.A., Rovine, M.J. (2023). Latent trajectories of learning-and teacher-context behavior problems across the primary school transition. Journal of Applied Developmental Psychology, 86: 101538. https://doi.org/10.1016/j.appdev.2023.101538

[7] Ward, K., Zelinsky, A. (2000). Acquiring mobile robot behaviors by learning trajectory velocities. Autonomous Robots, 9(2): 113-133. https://doi.org/10.1023/A:1008914200569

[8] Saravanan, A., Shade, M., Liu, Y., Olayeni, B., Sanders, S., Johnson, R., Starkweather, A. (2024). Training to use smart tablets to access reliable online health information in older adults' post-pandemic: A focused pilot intervention study. Geriatric Nursing, 56: 204-211. https://doi.org/10.1016/j.gerinurse.2024.02.010

[9] Carter, C.L., Carter, R.L., Foss, A.H. (2018). The flipped classroom in a terminal college mathematics course for liberal arts students. Aera Open, 4(1): 2332858418759266. https://doi.org/10.1177/2332858418759266

[10] Magyari, L., De Ruiter, J.P., Levinson, S.C. (2017). Temporal preparation for speaking in question-answer sequences. Frontiers in Psychology, 8: 211. https://doi.org/10.3389/fpsyg.2017.00211

[11] Li, Y., Guo, X., Lin, W., Zhong, M., Li, Q., Liu, Z., Zhu, Z. (2021). Learning dynamic user interest sequence in knowledge graphs for click-through rate prediction. IEEE Transactions on Knowledge and Data Engineering, 35(1): 647-657. https://doi.org/10.1109/TKDE.2021.3073717

[12] Wang, Y., Yang, Y. (2025). Efficient visual transformer by learnable token merging. IEEE Transactions on Pattern Analysis and Machine Intelligence, 47(11): 9597-9608. https://doi.org/10.1109/TPAMI.2025.3588186

[13] Bedair, S.S., Pulskamp, J.S., Rudy, R., Polcawich, R., Cable, R., Griffin, L. (2018). Boosting MEMS piezoelectric transformer figures of merit via architecture optimization. IEEE Electron Device Letters, 39(3): 428-431. https://doi.org/10.1109/LED.2018.2799864

[14] Zhou, Q., Sheng, K., Zheng, X., Li, K., Tian, Y., Chen, J., Ji, R. (2024). Training-free transformer architecture search with zero-cost proxy guided evolution. IEEE Transactions on Pattern Analysis and Machine Intelligence, 46(10): 6525-6541. https://doi.org/10.1109/TPAMI.2024.3378781

[15] Jiménez-Navarro, M.J., Martínez-Ballesteros, M., Martínez-Álvarez, F., Asencio-Cortés, G. (2023). A new deep learning architecture with inductive bias balance for transformer oil temperature forecasting. Journal of Big Data, 10(1): 80. https://doi.org/10.1186/s40537-023-00745-0

[16] Maresca, D., Correia, M., Tanter, M., Ghaleh, B., Pernot, M. (2018). Adaptive spatiotemporal filtering for coronary ultrafast Doppler angiography. IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control, 65(11): 2201-2204. https://doi.org/10.1109/TUFFC.2018.2870083

[17] Wang, Y., Jing, C. (2022). Spatiotemporal graph convolutional network for multi-scale traffic forecasting. ISPRS International Journal of Geo-Information, 11(2): 102. https://doi.org/10.3390/ijgi11020102

[18] Ma, W., Sartipi, K., Bender, D. (2016). Knowledge-driven user behavior pattern discovery for system security enhancement. International Journal of Software Engineering and Knowledge Engineering, 26(3): 379-404. https://doi.org/10.1142/S0218194016500169

[19] Huaulmé, A., Voros, S., Riffaud, L., Forestier, G., Moreau-Gaudry, A., Jannin, P. (2017). Distinguishing surgical behavior by sequential pattern discovery. Journal of Biomedical Informatics, 67: 34-41. https://doi.org/10.1016/j.jbi.2017.02.001