Proceedings of the 7th International Conference on Computing and Data Science
DOI: 10.54254/2755-2721/2025.26109

Intelligent Classification Model for Animation Style Based on Machine Learning Algorithm

Xinrui Peng[1], Xinghan Tian[1*]

[1]*School of Digital Technology & Innovation Design, Jiangnan University, Wuxi, China*
**Corresponding Author. Email: 272636093@qq.com*

Abstract. This study aims to improve the accuracy of animation style classification and provide more effective machine learning algorithm support for the intelligent retrieval, personalized recommendation, and intelligent development of creative assistance tools on animation content platforms. This paper innovatively proposes an optimization model (BiLSTM Transformer) that integrates bidirectional long short-term memory network (BiLSTM) and Transformer architecture to address the complex sequence features contained in animation data. In order to comprehensively evaluate the performance of the model, we conducted systematic comparative experiments with various representative models such as random forest, decision tree, XGBoost, CatBoost, and BP neural network. The experimental results show that the proposed BiLSTM Transformer model has achieved a significant breakthrough of 95.7% in classification accuracy, far exceeding the performance of the suboptimal model (91.8%), demonstrating a performance advantage of a discontinuous approach. Moreover, the model has consistently achieved over 95% accuracy, recall, and F1 score in key evaluation metrics, demonstrating its outstanding comprehensive performance and robustness. In contrast, the accuracy of all compared models is below 92% and shows a stepwise downward trend: the decision tree model (81.2%) has obvious overfitting problems; Random forest, as the best traditional method, still lags behind the new model by 3.9 percentage points in accuracy (91.8%); XGBoost and CatBoost are limited in their effectiveness due to the difficulty of fully learning the complex dependencies between sequences; BP neural network performs the weakest in nonlinear modeling ability due to the vanishing gradient problem. These comparative results fully verify that the BiLSTM Transformer model can effectively model and understand complex sequence patterns and style features in animation data by combining the powerful capturing ability of BiLSTM for long-distance contextual information and the focusing advantage of Transformer's self attention mechanism on key features. As a result, it has achieved excellent predictive performance in animation style classification tasks, providing strong technical support for related intelligent applications.

Keywords: Machine learning, intelligent classification of animation styles, Transformer

DOI: 10.54254/2755-2721/2025.26109

1. Introduction

The booming development of the animation industry has brought about a massive and increasingly diverse range of works, from traditional hand drawn 2D animations to intricate 3D CG, from the unique aesthetics of Japanese animation to realistic or exaggerated expressions in European and American animation, and even avant-garde explorations in experimental animation [1]. The style dimensions are extremely rich. Faced with such a huge amount of data and a complex style system, traditional classification methods that rely on expert manual annotation and subjective experience judgment are becoming increasingly inefficient, costly, and difficult to ensure objectivity and consistency [2]. Especially in the era of digital media, animation content is growing exponentially, and platforms have an urgent need for efficient content management, precise personalized recommendations (such as pushing specific style animations based on user preferences), and in-depth market style trend analysis [3]. At the same time, academic research urgently needs more systematic and quantifiable tools to analyze the evolution patterns, cross-cultural differences, and creator imprints of animation art styles [4]. Therefore, exploring automated and intelligent methods for classifying animation styles, establishing a scientifically rigorous classification system and efficient recognition technology, has become a key bridge connecting animation art research, industrial applications, and audience experience, with important theoretical and practical value.

Machine learning algorithms, especially deep learning techniques, provide powerful solutions to address the aforementioned challenges. Its core advantage lies in the ability to automatically learn and extract hidden, high-level style feature patterns from massive animation frame sequences, character design, scene composition, color usage, motion patterns, and other low-level visual and dynamic data. These patterns are often complex and subtle, even beyond the intuitive recognition ability of the human eye. Convolutional neural networks (CNNs) can effectively capture static style elements such as textures, lines, and color distribution in images; Recurrent neural networks (RNNs) or 3D convolutional networks (3D CNNs) are adept at analyzing dynamic style traits in temporal dimensions such as camera movements and character action rhythms; Traditional algorithms such as Support Vector Machine (SVM) and Random Forest can also achieve efficient classification when combined with effective feature engineering. Transfer learning utilizes pre trained models on large-scale image datasets such as ImageNet, significantly alleviating the problem of relatively scarce animation style annotation data and improving the model's generalization ability in small sample sizes [5]. Unsupervised or semi supervised learning can help explore and discover potential new style clusters or sub genres. These algorithms together promote the classification of animation styles from relying on subjective experience to being driven by objective data, greatly improving the accuracy, efficiency, and scalability of classification. They lay a solid technical foundation for intelligent retrieval and recommendation of animation content platforms, intelligent creation assistance tools, academic quantification research on animation art styles, and digital style archiving of cultural heritage, profoundly transforming the organization, understanding, and application of animation content. This article optimizes the Transformer model based on bidirectional long short-term memory networks for the classification of animation styles, providing corresponding machine learning algorithms for intelligent retrieval and recommendation of animation content platforms and intelligent creation assistance tools.

2. Data sources

This dataset contains 581 animation samples, covering four main styles: traditional hand drawn, 3D computer animation, stop motion animation, and Japanese animation. Each sample has 10

DOI: 10.54254/2755-2721/2025.26109

continuous numerical feature variables (including color diversity, line complexity, texture depth, motion fluency, shadow intensity, perspective depth, detail density, saturation level, contrast ratio, and stylization index). This dataset reflects the visual feature distribution of real animation works and aims to provide basic training and testing data for animation style classification algorithm research. Select some data for display, and the results are shown in Table 1.

Table 1. Selected partial dataset

Color_Variety	Line_Complexity	Texture_Depth	Motion_Fluidity	Shadow_Intensity	Perspective_Depth	Contrast_Ratio	Stylization_Index	Animation_Style
2.43	-4.43	-10.29	9.2	6.95	4.56	3.78	-2.32	Anime
2.46	-7.21	-8.63	7.12	8.51	5.87	3.97	0.21	3D CG
1.99	-10.09	-10.04	6.89	11.24	7.05	3.07	1.24	Traditional
-11.08	10.14	8.6	-2.74	-5.61	-5.1	-3.68	-2.53	Anime
-5.63	8.79	3.66	2.12	-6.94	-7.46	3.29	5.27	Stop Motion
1.66	-9.96	-5.27	-5.13	-1.06	9.08	0.2	-7.93	3D CG
-2.15	6.91	5.83	2.38	-11.45	-6.94	4.04	5.4	Stop Motion
-4.09	9.83	3.66	0.26	-6.91	-8.08	4.89	4.23	3D CG
-10.32	9.68	6.96	-4.14	-6.92	-4.73	-0.23	-1.96	Stop Motion
3.77	-9.12	-9.72	8.76	10.84	5.69	2.66	-2.16	Anime
-9.8	10.27	7.73	-5	-6.23	-6.62	-2.22	-2.76	Anime
-8.84	9	8	-1.65	-3.68	-7.98	-2.18	-3.38	3D CG

3. Method

3.1. BiLSTM

Bidirectional Long Short Term Memory Network (BiLSTM) is an improved recurrent neural network structure. The network structure of BiLSTM is shown in Figure 1, and its core idea is to simultaneously utilize the forward and backward information of sequence data in time [7]. It consists of two independent LSTM layers: one processes the input sequence in chronological order (from start to finish), capturing the dependence of historical information on the current time; The other one processes the same input sequence in reverse order (from the end of the sequence to the beginning), capturing the impact of future information on the current moment. The hidden states obtained from these two directions are connected or merged at each time step, allowing the network to synthesize rich information from the complete context of the sequence at any given time [8].

DOI: 10.54254/2755-2721/2025.26109

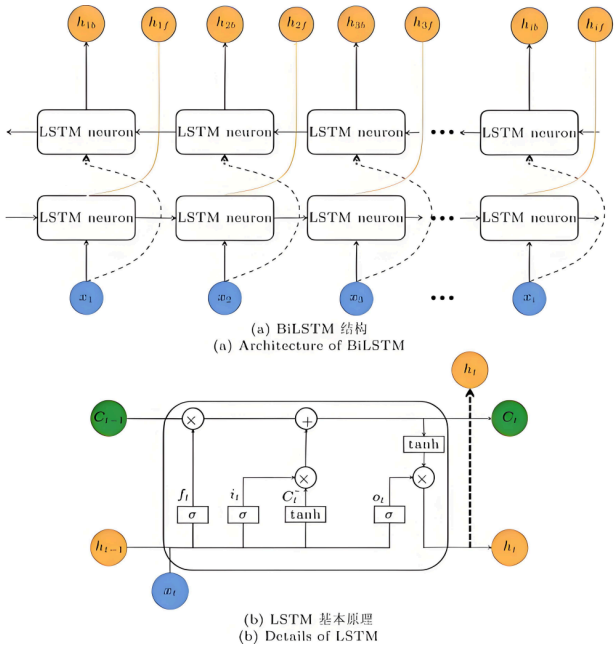

Figure 1. The network structure of BiLSTM

This bidirectional structure greatly enhances the model's ability to understand sequential contexts. Each LSTM unit in each direction is controlled by a precise gating mechanism (input gate, forget gate, output gate) to regulate the flow of information, which can learn long-range dependencies and effectively alleviate the gradient vanishing or exploding problem faced by traditional RNNs.

3.2. Transformer

Transformer is a deep learning architecture based on self attention mechanism, which has completely changed the field of natural language processing. The network structure of Transformer is shown in Figure 2, and its core lies in abandoning traditional cyclic or convolutional structures and relying entirely on attention mechanisms to model global dependencies in sequence data [9]. The Transformer model consists of a stack of encoders and decoders. Encoder processing of input sequences (such as source language sentences): Firstly, the input lexical elements are transformed into vectors through word embedding layers and overlaid with positional encoding to inject sequence order information; Subsequently, these vectors are input into multiple identical encoder layers. Each encoder layer consists of two key sub layers: a multi head self attention layer and a feedforward neural network layer. The self attention mechanism dynamically generates a representation containing the global context for each word element by calculating the association weight (called attention score) between each word element in the sequence and all other word elements. Multi head design allows the model to focus on information from different representation subspaces in parallel. Each sub layer uses residual connections and layer normalization to stabilize training. The decoder works in a similar but more complex way, gradually generating an output sequence [10]. It also includes embedding layers, positional encoding, and multiple identical decoder layers. Each decoder layer has three sub layers: the first is a masked multi head self attention layer; The second one is the encoder decoder attention layer; The third one is the feedforward neural network layer. The decoder works through autoregression, receiving the previously generated output as input at each step, and ultimately predicting the probability distribution of the next morpheme through a linear layer and softmax function. This parallelized self

DOI: 10.54254/2755-2721/2025.26109

attention mechanism enables Transformers to efficiently capture long-range dependencies, becoming the foundation of revolutionary models such as BERT and GPT.

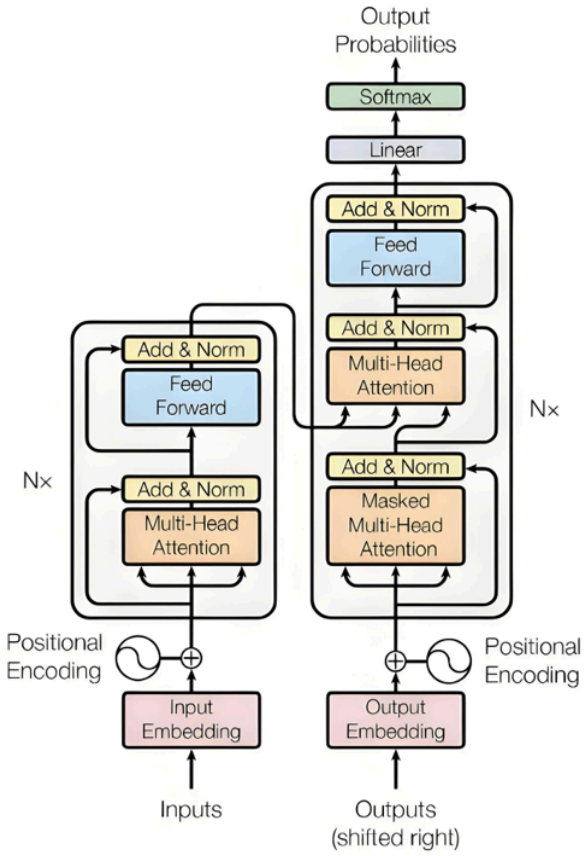

Figure 2. The network structure of Transformer

3.3. BiLSTM-Transformer

The core principle of BiLSTM (Bidirectional Long Short Term Memory) optimization for Transformer lies in utilizing its powerful temporal modeling capability to compensate for the potential shortcomings of Transformer in capturing local dependencies and explicit sequence order information. Transformers rely on self attention mechanisms to establish global dependencies, but their positional encoding may not be precise or efficient enough for modeling sequence order, especially when dealing with long sequences, where the capture of local continuous patterns may be weaker than that of cyclic structures. BiLSTM, through its unique gating mechanism and cellular state, is able to explicitly learn complex long-term and short-term dependencies between elements in a sequence. Its bidirectional nature enables it to more fully understand the meaning of each element in the context of the complete sequence. The hidden state vectors learned by BiLSTM, which contain fine temporal dynamics and local context, provide better input features or supplementary features for the self attention mechanism of Transformer, enabling the model to more accurately associate elements with strong temporal correlation when calculating attention weights, thereby improving the performance of Transformer in tasks that require precise understanding of word order and local structure, and ultimately forming a hybrid architecture that combines global attention and local loop advantages.

DOI: 10.54254/2755-2721/2025.26109

4. Result

The hardware adopts dual NVIDIA Tesla A100 80GB GPU (with a total of 160GB of video memory), AMD EPYC 7763 64 core processor, and 256GB DDR4 memory. The software environment is Matlab R2024a; The model adopts a 12 layer Transformer encoder (hidden layer dimension 768, 12 head attention) followed by a bidirectional LSTM layer (hidden unit 512), with a word embedding dimension of 256. The training uses an AdamW optimizer (learning rate 5e-5, weight decay 0.01), cosine annealing scheduling (maximum cycle 20), early stopping mechanism (patience value 5 rounds), batch size 64, Dropout rate 0.3, and maximum sequence length 512.

This article uses random forest, decision tree, XGBoost, CatBoost, and BP neural network as comparative models, and evaluates the classification performance of the models using Accuracy, Precision, Recall, and F1. The results of the comparative experiments are shown in Table 2. The bar chart comparison of the comparative experiment is shown in Figure 3.

Table 2. The results of the comparative experiments

Model	Accuracy (%)	Precision (%)	Recall (%)	F1 (%)
BP Neural Network	85.4	84.7	86.1	85.4
Random Forest	91.8	92.1	91.5	91.8
Decision Tree	81.2	80.5	81.9	81.2
CatBoost	88.3	87.9	88.6	88.2
XGBoost	89.6	90.3	88.9	89.6
BiLSTM-Transformer	95.7	96.2	95.3	95.7

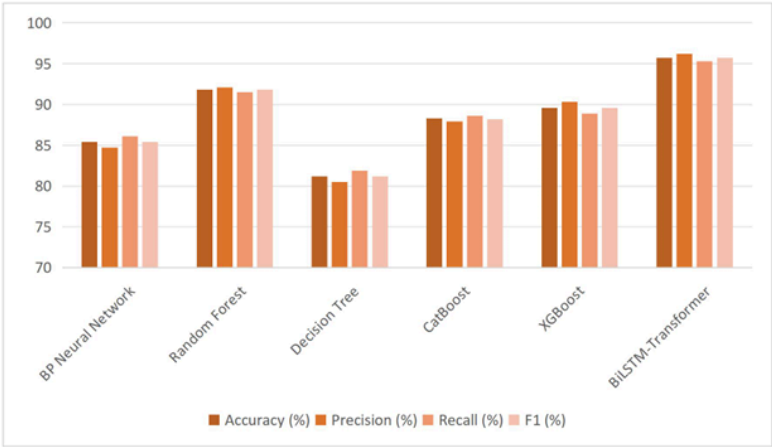

Figure 3. The bar chart comparison of the comparative experiment

The experimental results show that the BiLSTM Transformer model exhibits significant performance advantages, with an accuracy of 95.7% far exceeding the highest performance of other models, which is 91.8%, forming a fault zone leadership; At the same time, the model consistently exceeded 95% in accuracy, recall, and F1 score, while the comparison model not only had accuracy rates all below 92%, but also showed a step like decline. Among them, the decision tree model only had an accuracy of 81.2%, exposing serious overfitting defects. Although the random forest method was the best among traditional methods, it still lagged behind by 3.9 percentage points. XGBoost and CatBoost were limited in their effectiveness due to the difficulty of fully learning sequence

DOI: 10.54254/2755-2721/2025.26109

dependencies, while BP neural network performed the weakest in nonlinear modeling due to gradient vanishing problems. This fully verifies that the design of BiLSTM Transformer fusion bidirectional long short-term memory network and self attention mechanism can effectively capture complex sequence features and demonstrate excellent performance in animation style prediction.

5. Conclusion

This article delves into the key task of animation style classification and innovatively proposes and validates an optimization model (BiLSTM Transformer) that integrates bidirectional long short-term memory network (BiLSTM) and Transformer architecture. Through rigorous experimental design, this study comprehensively compared the proposed model with classic machine learning and deep learning models such as random forest, decision tree, XGBoost, CatBoost, and BP neural network. The experimental results clearly and strongly demonstrate that the BiLSTM Transformer model exhibits a discontinuous performance advantage in animation style classification tasks. Its core advantage lies in an astonishing 95.7% classification accuracy, which not only significantly surpasses the best comparison model (91.8%) by 3.9 percentage points, but also highlights the outstanding discriminative ability of the model. More importantly, the model consistently surpasses the 95% threshold in core evaluation metrics such as accuracy, recall, and F1 score, demonstrating high robustness and balance.

In contrast, all compared models not only have accuracy rates below 92%, but also exhibit a stepwise decline in performance. The decision tree model only has an accuracy of 81.2%, exposing serious overfitting issues; Random forest, as the best representative of traditional ensemble methods, still lags significantly behind the BiLSTM Transformer model in terms of performance; The gradient boosting tree models such as XGBoost and CatBoost are inadequate in dealing with the inherent dependencies of sequential data, resulting in limited effectiveness; However, BP neural network performs the weakest in capturing highly nonlinear animation style features due to the vanishing gradient problem. This series of comparative results fully validates the effectiveness of the BiLSTM Transformer model design: by organically combining the advantages of BiLSTM in capturing long-range contextual temporal dependencies with the powerful ability of Transformer self attention mechanism in global feature modeling and relationship extraction, the model can efficiently and deeply learn complex sequence patterns and deep semantic information contained in animation data. The most important significance of this study is that the developed BiLSTM Transformer model provides a high-performance and highly reliable core algorithm engine for the intelligent retrieval and precise recommendation system of animation content platforms, as well as the intelligent upgrade of animation creation auxiliary tools, which effectively promotes the intelligent development process of the animation industry in the era of artificial intelligence. The algorithm proposed in this article significantly improves the accuracy, efficiency, and scalability of classification, laying a solid technical foundation for intelligent retrieval and recommendation of animation content platforms, intelligent creation assistance tools, academic quantification research on animation art styles, and digital style archiving of cultural heritage.

References

[1] He, Jia. "Exploring style transfer algorithms in Animation: Enhancing visual." Entertainment Computing 49 (2024): 100625.

[2] Zheng, Jie. "Design and Application of Intelligent Processing Technology for Animation Images Based on Deep Learning." Mobile Information Systems 2022.1 (2022): 9438086.

DOI: 10.54254/2755-2721/2025.26109

[3] Kan, Mengyang, and Ziyun Liu. "Design of Color Matching Algorithm for Animation Character Background Based on Decision Tree Classification Model." 2023 Asia-Europe Conference on Electronics, Data Processing and Informatics (ACEDPI). IEEE Computer Society, 2023.
[4] Wang, Qingqing. "Research on the Training Mode of Animation Professionals in the Era of Artificial Intelligence." Advances in Education, Humanities and Social Science Research 13.1 (2025): 444-444.
[5] Shank, Daniel B., et al. "AI composer bias: Listeners like music less when they think it was composed by an AI." Journal of Experimental Psychology: Applied 29.3 (2023): 676.
[6] Yang, Tiancheng, and Shah Nazir. "A comprehensive overview of AI-enabled music classification and its influence in games." Soft Computing 26.16 (2022): 7679-7693.
[7] Hu, Bofei. "Distributed Sensing of Animation Art Style under Big Data Technology." 2021 3rd International Conference on Artificial Intelligence and Advanced Manufacture. 2021.
[8] Li, Shan. "Application of artificial intelligence-based style transfer algorithm in animation special effects design." Open Computer Science 13.1 (2023): 20220255.
[9] Guo, Nili. "Feature Extraction Method for Shot Based Animation Script Creation Empowered by Artificial Intelligence." International Journal of High Speed Electronics and Systems (2025): 2540352.
[10] Abdulraheem, Ayah Wafiq, Mohammad Arafah, and Ahmad Aladawi. "Intelligent Age Group Classification System for Animated Films Using Hybrid Machine Learning Models." 2025 1st International Conference on Computational Intelligence Approaches and Applications (ICCIAA). IEEE, 2025.