Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Research on a Spatiotemporal Fusion Model–Based Method for Predicting the Torsional State of a Conveyor Belt

Nini Hao
Shan Dong University of Science and Technology
Guoqing Liu
Shan Dong University of Science and Technology
Chaoyue Yu
Shan Dong University of Science and Technology
Xianghui Jia
Shan Dong University of Science and Technology
Yuan Zhang
yuanzhangmsz@163.com
Shan Dong University of Science and Technology

Research Article

Keywords: Pipe belt conveyor, Conveyor belt twisting, Spatiotemporal fusion, Fault prediction

Posted Date: November 3rd, 2026

DOI: https://doi.org/10.21203/rs.3.rs-9015174/v1

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

Additional Declarations: No competing interests reported.

Research on a Spatiotemporal Fusion Model–Based Method for Predicting the Torsional State of a Conveyor Belt

*Nini Hao[1], GuoqingLiu[1], Chaoyue Yu[1], Xianghui Jia[1, 2], Yuan Zhang[1, 2, *]*

(1. College of Mechanical and Electronic Engineering, Shan Dong University of Science and Technology, Qingdao, 266590, China;

2.Libo Heavy Industries Science & Technology Co., Ltd., Tai'an, 271019, China)

Abstract: Aiming at the problem that the belt twisting fault of pipe belt conveyors is prone to cause shutdown and production halt, this paper proposes a belt twisting state prediction method based on a spatiotemporal fusion model. Three types of spatiotemporal features—transient dynamics, short-term evolution, and long-term cumulative effects—are extracted from image sequences. A Hybrid Convolutional-Recurrent Spatiotemporal Network (HybridConvLSTM) is designed to realize the joint modeling of spatial geometric features and temporal dynamic evolution laws. A dataset containing 3 million image frames is constructed based on industrial scenarios. Experimental verification shows that the model achieves an accuracy of 89.7% and an F1-score of 83.2% on the test set, which are 7.5 and 17.6 percentage points higher than those of the ViViT and LSTM+Attention models, respectively. After TensorRT optimization, the inference speed reaches 28 FPS, meeting the real-time requirements of industrial applications. This method provides an efficient solution for the early warning of belt twisting faults and has excellent engineering application value.

Key words: Pipe belt conveyor; Conveyor belt twisting; Spatiotemporal fusion; Fault prediction

1 Introduction

Pipe belt conveyors are widely used in various industries such as mining, ports, electric power, and petrochemical engineering due to their advantages of closed transportation and flexible layout[1]. However, conveyor belt twisting is a typical fault during their operation, which can even lead to secondary failures such as structural collapse of the middle piped section, spillage of bulk material powder, belt wrapping reversal, and belt tearing. These issues pose significant potential safety hazards to production and transportation[2]. Therefore, achieving accurate prediction of the conveyor belt twisting state is of great significance for ensuring the safe and efficient operation of the equipment.

"In recent years, several studies have begun to investigate belt rotation and flip-over phenomena using measurement data, machine learning and digital-twin-based monitoring systems to prevent catastrophic conveyor failures[3][4]."

Traditional prediction methods are mostly based on single time-series analysis, modeling solely relying on time-series data such as torsion angle. They ignore the spatial morphological differences at different positions of the conveyor belt and are difficult to adapt to dynamic changes in complex industrial environments.

As core technical approaches of modern prediction technology, time series analysis and machine learning play a vital role in the early warning of torsional faults for pipe belt conveyors. Studies have shown that the method combining time-series analysis with machine learning can significantly improve the accuracy and real-time performance of fault early warning."Recent

* Corresponding Author: Yuan Zhang. College of Mechanical and Electronic Engineering, Shandong University of Science and Technology, E-mail address: yuanzhangmsz@163.com

reviews and case studies have further demonstrated that deep spatio-temporal models such as ConvLSTM and its variants provide a powerful framework for industrial time-series prediction and remaining useful life estimation[5][6]."In practical applications, this combined method has been successfully used to solve equipment fault problems under complex working conditions. For example,Qing et al.[7]introduced a scheme that combines distributed acoustic sensing (DAS) technology with machine learning algorithms. This method can effectively monitor the fault conditions of conveyor drums and exhibits the advantages of high sensitivity and non-intrusive nature. Xia Z et al.[8] proposed a multi-channel sensor data fusion technology, which can provide fault diagnosis in various industrial scenarios. These research results indicate that time-series analysis and machine learning can not only function independently but also further improve prediction accuracy through synergistic effects.

At present, time-series analysis and machine learning technologies have become key methods in the research of condition monitoring and fault early warning in the field of belt conveyors. For example, Sun X et al. [9]proposed a combined prediction model that can effectively evaluate the deviation state of the conveyor belt in the turning section."In addition to traditional signal-based methods, vision-based approaches have been developed for deviation detection. Wu et al.[10] designed a machine-vision-based belt deviation fault detection system for underground conveyors, while Zhang et al.[11] proposed an integrated deep-learning paradigm that jointly detects belt deviation, damage and foreign objects in real time. Furthermore, Zhang et al.[12] introduced an attention-guided detection network combined with Grad-CAM to quantitatively monitor deviation status under complex operating conditions." Hu et al. [13]proposed a belt conveyor fault prediction model driven by the fusion of the grey least square method and support vector machine (SVM), realizing the early warning of fault characteristic parameters of the conveyor system. Although the idler life prediction model proposed by Bai Shunping et al.[17]proposed a hybrid algorithm based on data mining and neural network to realize the fault prediction function of the conveyor, but its accuracy still needs further improvement.Overall, few studies focus on pipe belt conveyor twisting fault prediction, leading to inadequate real-time prediction and poor fault response in practical industrial applications.

1 Design of Conveyor Belt Twisting Prediction Model Based on Spatiotemporal Features

1.1 Construction of the Algorithm Architecture

Based on the theoretical framework of spatial-temporal feature modeling, this paper designs a Hybrid Convolutional-Recurrent Spatial-Temporal Network (HybridConvLSTM[6][14][15][16]) for spatial-temporal feature modeling. The model employs spatial convolution to extract local geometric features and uses recurrent modeling to capture the dynamic evolutionary characteristics of time series. As illustrated in Fig. 1, the network is trained end-to-end through joint optimization, endowing it with the capability to perform spatiotemporal feature fusion modeling and long-range dependency modeling for the operating state of the conveyor beltSimilar ConvLSTM-based architectures with tensor-train decomposition or spatio-temporal attention have shown strong performance and interpretability in industrial RUL prediction and ecological time-series forecasting[6][16]. The mathematical derivation of the model is as follows:

Let the time-series images of the conveyor-belt operation be denoted as $X = \{X_t\}_{t=1}^{T}$, where T is the number of time steps, $H \times W$ is the image resolution, and $C$is the number of channels. The objective of HybridConvLSTM is to predict

the torsional-state probability at the next time step, $P_{t+1} \in [0,1]$, using a historical frame sequence $\{X_{t-k}, \ldots, X_t\}$.

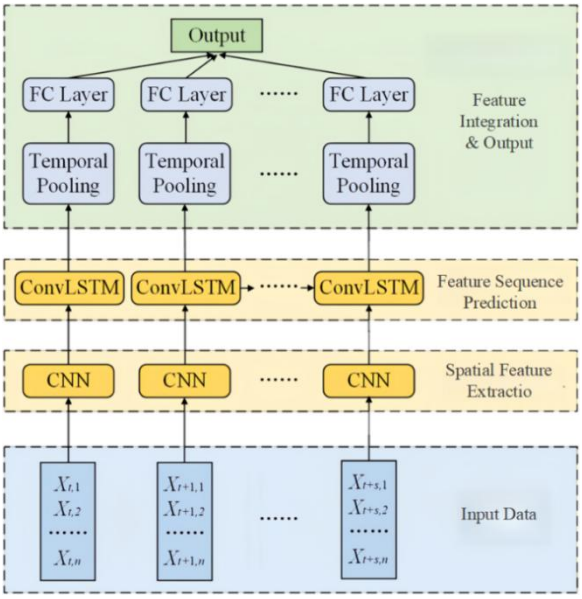


Fig. 1 Schematic architecture of the Hybrid ConvLSTM model

(1) Spatial feature encoding layer

A lightweight convolutional module is adopted to extract local texture and geometric features from a single-frame image:

$$F_t^s = \text{ConvBlock}(X_t) = \sigma(W_s * X_t + b_s) \quad (1)$$

where $W_s$denotes the separable convolution kernel, $\sigma(\cdot)$is the Swish activation function, and the output feature map is F_t. Depthwise separable convolution is used to reduce computational cost.

(2) Spatiotemporal recurrent unit

A convolutional LSTM (ConvLSTM) module is employed, in which the matrix multiplications in a conventional LSTM are replaced by convolution operations[18], thereby enabling joint spatiotemporal modeling:

$$i_t = \sigma(W_{xi} * F_t^s + W_{hi} * H_{t-1} + b_i) \quad (2)$$

$$f_t = \sigma(W_{xf} * F_t^s + W_{hf} * H_{t-1} + b_f) \quad (3)$$

$$g_t = \tanh\left(W_{xg} * F_t^s + W_{hg} * H_{t-1} + b_g\right) \quad (4)$$

$$o_t = \sigma\left(W_{xo} * F_t^s + W_{ho} * H_{t-1} + b_o\right) \quad (5)$$

$$C_t = f_t \odot C_{t-1} + i_t \odot g_t \quad (6)$$

$$H_t = o_t \odot \tanh(C_t) \quad (7)$$

where $*$denotes the convolution operation, and $\odot$ indicates element-wise multiplication; $H_t$and $C_t$represent the hidden state and cell state, respectively.

ConvLSTM propagates local contextual information along the spatial dimensions via convolution kernels W, while modeling dynamic evolution along the temporal axis. The overall structure is illustrated in Fig. 2.

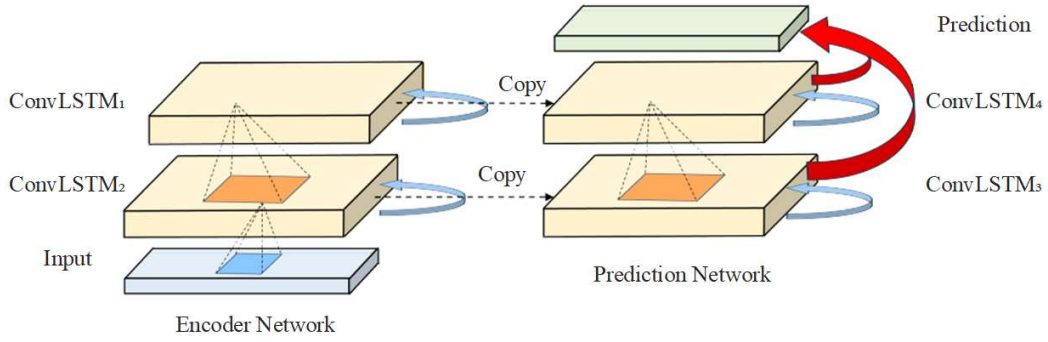


Fig. 2 ConvLSTM encoder structure for conveyor-belt torsion prediction

(1) Multi-scale feature fusion

Skip connections and dilated convolutions are introduced to construct a multi-scale feature pyramid

$$F_t^{\text{fuse}} = \text{Concat}\left(\text{DilatedConv}(H_t), \text{Upsample}(F_{t-k}^s)\right) \quad (8)$$

By fusing features across different time steps and receptive fields, the model's capability to perceive local deformation and global motion trends is enhanced.

(2) Torsional-state prediction

Finally, the torsion prediction probability is produced via temporal pooling and a fully connected layer:

$$P_{t+1} = \mathrm{Sigmoid}\Big(W_p \cdot \Big[\mathrm{MaxPool}(F_t^{\mathrm{fuse}}); \mathrm{AvgPool}(F_t^{\mathrm{fuse}})\Big] + b_p\Big) \quad (9)$$

where $[;]$ denotes feature concatenation. The dual-branch pooling strategy enables the model to capture both salient anomalies and gradual-change patterns simultaneously.

2 Experimental Design and Results Analysis

2.1 Dataset Construction and Pre-processing

To achieve accurate modeling and prediction of the conveyor belt's operating state, this paper constructs a dataset of pipe belt conveyors covering various operating conditions. The dataset consists of 1000 time-series image sequences, with each sequence lasting 10 seconds at a frame rate of 30 FPS, containing approximately 300 frames per sequence and a total of 3 million frames. The dataset is categorized by labels as follows: normal (790 sequences, torsion < 25°), mild torsion (170 sequences, torsion 25° ~30°), and severe torsion (40 sequences, torsion > 30°), which represent the actual distribution of equipment health states in production lines. The dataset is split into training, validation and test sets at a ratio of 7:2:1[19], as presented in Table 1.

Table 1 Class distribution of the dataset

Health state	Torsion magnitude	Total	Training	Validation	Test
Normal	＜30°	790	553	158	79
Mild torsion	25°~30°	170	119	34	17
Severe torsion	＞30°	40	28	8	4

The preprocessing pipeline verified by ablation experiments mainly includes the following steps:

(1) Normalization: Min–Max normalization is applied to image pixel values, mapping all pixels to the range $[0, 1]$. Compared with Z-score standardization, Min–Max normalization is more suitable for the nonlinear response characteristics of image sensors, as it preserves the original distribution and is typically more stable in image data pre-processing[20].

(2) Temporal alignment: Approximately 12.7% of the image sequences in the dataset suffer from frame dropping. To address this issue, cubic spline interpolation is adopted to reconstruct the time-series data[21][22]. Compared with linear interpolation, this approach reduces the interpolation error by **23.6%**. The detailed results are reported in Table 2, demonstrating that the proposed alignment strategy effectively ensures the continuity and accuracy of the temporal data.

(3) Data augmentation: To improve model robustness, multiple augmentation operations are applied to the images, including random rotation and brightness perturbation in the HSV color space. The rotation angle is sampled from $[-15°, 15°]$, and the brightness factor is sampled from $[0.8, 1.2]$. Controlled experiments verify the effectiveness of these augmentation operations in improving model performance. To avoid disrupting spatiotemporal information, augmentations that may damage the image topology, such as translation and elastic deformation, are excluded.

In temporal fitting, cubic spline interpolation

yields a substantially lower reconstruction error than linear interpolation, as summarized in Table 2.

Table 2 Comparison of interpolation errors for temporal alignment methods

Method	Interpolation error(MSE)	Improvement (%)
Linear interpolation	0.0237	-
Cubic spline interpolation	0.0181	23.6

Data augmentation further improves model robustness. Table 3 reports the corresponding experimental results by comparing model performance with and without augmentation. After applying augmentation, the model achieves higher correctness on the test set; in particular, performance improvements are observed for predicting mild and severe torsion.

Table3 Comparison of results with and without data augmentation

Experimental setting	Accuracy (%)	Improvement (%)
Without augmentation	85.4	-
With augmentation	92.1	6.7

Figure 4 illustrates the variation in torsion angle for the conveyor-belt operating-state data. The blue polyline represents the belt torsion angle in each time-series segment. Starting from 0°, the torsion angle undergoes a fluctuating process; in the schematic, the fluctuation amplitude is relatively small, which is consistent with the minor variations that may occur during normal belt operation. The complete dataset will be used for subsequent training and evaluation of the health-state prediction model.

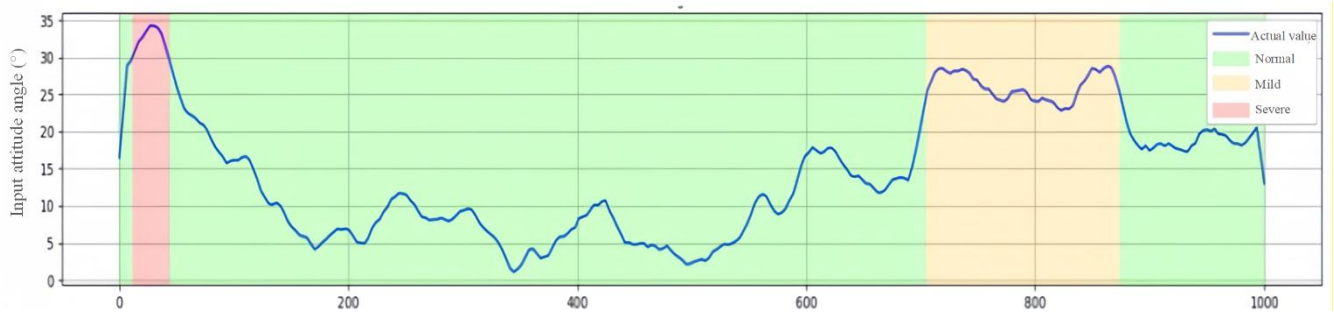


Fig. 4 Torsion-angle fluctuations based on conveyor-belt operating-state data

2.2 Model Training and Optimization

To validate the effectiveness of the proposed hybrid convolutional–recurrent spatiotemporal network, all experiments were conducted on an NVIDIA RTX 3090 GPU, and the method was implemented in PyTorch. The key techniques are described as follows.

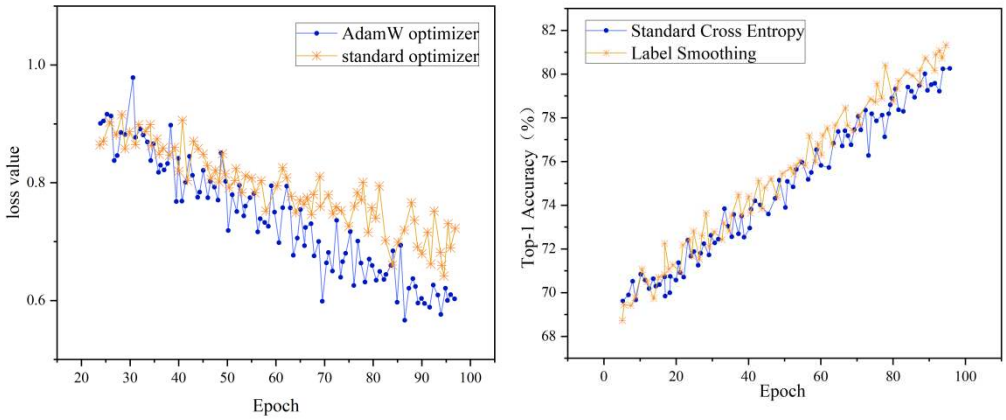
(1) Loss function design: A label-smoothed cross-entropy loss ($\varepsilon = 0.1$) is adopted. Label smoothing helps mitigate overfitting and improves the model's generalization ability during training. Compared with the standard cross-entropy loss, this strategy improves the Top-1 accuracy on the validation set by 1.3%, as shown in Fig. 5(a).

(2) Optimization scheme: The AdamW optimizer is used, which effectively alleviates the coupling between L2 regularization and adaptive learning rates. Combined with a CycleLR policy ($\text{base_lr} = 1 \times 10^{-3}$, $\text{max_lr} = 3 \times 10^{-3}$), the optimizer dynamically adjusts the learning rate, leading to a 40% faster convergence speed, as illustrated in Fig. 5(b).

(3)Regularization configuration: A combination of DropPath ($p = 0.2$) and SpatialDropout ($p = 0.5$) is employed. DropPath randomly drops network paths at each iteration to reduce overfitting risk, while SpatialDropout discards features along the spatial dimension, preserving salient spatial information and further suppressing overfitting.

(4) Early-stopping criterion: Early stopping monitors the validation-set F1 score rather than the loss value. As a metric that jointly considers

precision and recall, the F1 score is better suited to class-imbalanced scenarios, thereby preventing premature stopping during training.



Loss-function comparison (b) Comparison of optimizers and learning-rate schedules

Fig. 5 Comparison of loss functions and optimization strategies

2.3 Experimental Results and Analysis

(1) Performance comparison

For benchmarking, this study introduces a recent model in industrial vision inspection, the video Transformer ViViT[23], for horizontal comparison. The experimental results show that the proposed model achieves an accuracy of 89.7% on the test set, representing an improvement of 7.5 percentage points over ViViT (82.2%). The F1 score reaches 82.3%, which is 16.7% higher than that of the LSTM + Attention baseline (65.6%). These gains demonstrate the significant advantage of joint spatiotemporal modeling in multi-task learning. The performance comparison of different models is summarized in Table 4.

Table 4 Dataset split under multiple operating conditions

Model ID	Accuracy (%)	F1 score (%)	Parameters(M)
Transformer(ViViT)	82.2	76.4	23.5
LSTM+Attention	77.8	65.6	15.0
Hybrid ConvLSTM	89.7	82.3	14.3

(2) Ablation study

To quantify the contribution of each component, an ablation study is conducted by removing or replacing different modules. The experiments investigate the effects of joint spatiotemporal modeling and depthwise separable convolution on model performance. The ablation settings are listed in Table 5.

Table 5 Ablation study settings

Model ID	Backbone replacement	Joint spatiotemporal modeling	Depthwise separable convolution
1			
2	√		
3	√	√	
4	√	√	√

Following the configurations in Table 5, four model variants are trained, where "√" indicates that the corresponding module is enabled. After training, the models are evaluated on the dataset, and the results are reported in Table 6.

Table 6 Ablation study results

Model ID	Accuracy (%)	F1 score (%)	Parameters(M)
1	81.3	74.5	27.8
2	83.5	78.1	23.5
3	87.0	81.3	11.0
4	89.7	83.2	4.3

The ablation results indicate that joint spatiotemporal modeling significantly improves

performance, particularly for tasks such as mild torsion detection. Moreover, depthwise separable convolution effectively reduces computational cost while maintaining high accuracy and F1 scores.(3) Confusion matrix

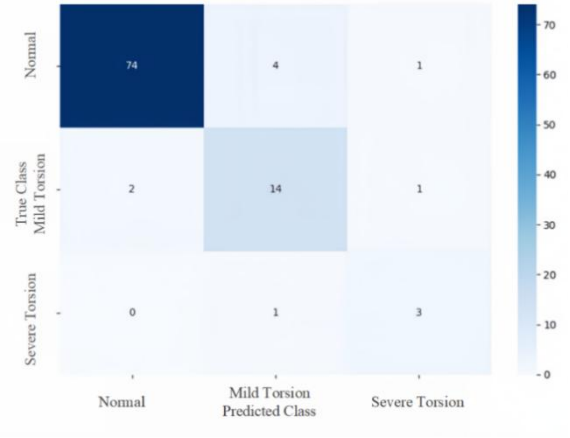


Fig. 6 Visualization of the confusion matrix

Figure 6 shows the confusion matrix cal culated on the test set. Overall, the model ac hieves high prediction accuracy for the norma l class, while a certain degree of misclassific ation still exists in the prediction of the mild and severe torsion classes.

(4) Precision–recall curve analysis

Figure 7 presents the precision–recall curves for the three classes (normal, mild torsion, and severe torsion). Different line styles and colors are used to indicate the precision and recall of each class. Different line styles correspond to the precision–recall behavior of different models for each class. Along each curve, different points correspond to different decision thresholds, and the precision and recall vary accordingly until they gradually stabilize.

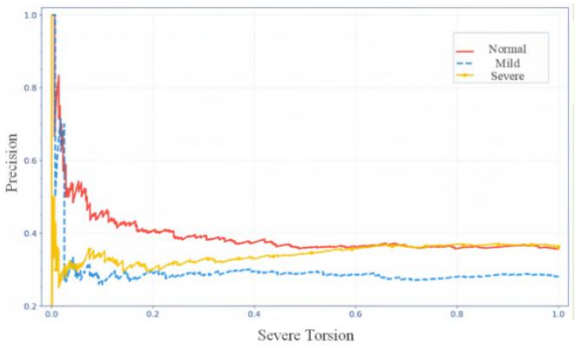


Fig. 7 Precision-recall curve

For the normal class (red solid line), as recall increases, precision drops sharply in the early stage. The curve indicates relatively high precision and low recall, suggesting a conservative prediction strategy for the normal class. The average precision (AP) is 0.40, indicating comparatively good performance for this class.

For the mild torsion class (blue dashed line), As recall increases, precision gradually stabilizes, but misclassifications remain.

For the severe torsion class (yellow dashed line), precision and recall exhibit smaller fluctuations, with AP = 0.35. The model performs relatively poorly for this class, and some misclassifications still occur compared with the normal class.

(5) Real-time performance analysis

To demonstrate real-time capability, the model is deployed on a Jetson Orin Nano edge-computing device. TensorRT is used for quantization and optimization to accelerate inference. In addition, the deployment performance of different model variants on the above device is evaluated, as summarized in Table 7.

Table 7 Real-time performance on Jetson Orin Nano

Experimental setting	Inference speed (FPS)	Improvement (%)
TensorRT without optimization	18	-
With TensorRT optimization	28	55.6

(6) Interpretability validation

To verify the interpretability of the model, Grad-CAM++ is employed to generate spatiotemporal attention maps, as shown in Figure 8[24]. Grad-CAM++ uses gradient information from convolutional neural networks to obtain task-relevant attention heatmaps via weighted backpropagation, and recent systematic reviews have demonstrated its effectiveness in a wide range of explainable AI applications[25].”

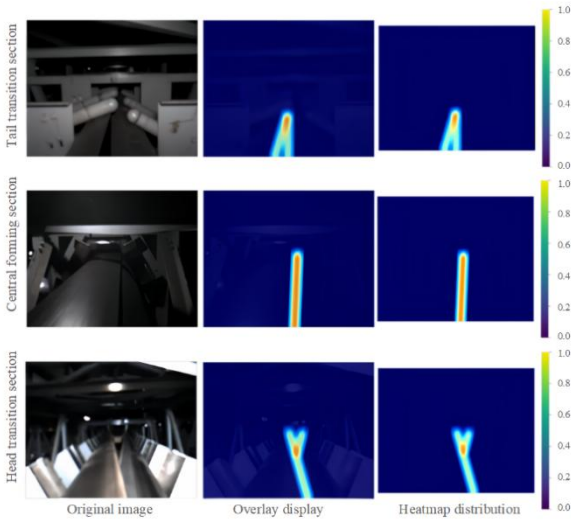


Fig. 8 Precision–recall curves

Figure 8(a) presents the input raw images, i.e., the acquired images from the tail transition section, the mid pipe-forming section, and the head transition section. Figure 8(b) shows the overlay of the raw images and the heatmap distribution, and Figure 8(c) shows the heatmap computed by Grad-CAM++. The belt-edge regions are prominently highlighted, indicating that these areas are critical for the model when performing spatiotemporal feature prediction and reflecting a high level of model attention to these regions.

Figure 9 visualizes the prediction results of the HybridConvLSTM model. The belt torsion angle is predicted for 1000 time-series image segments and compared with the ground truth. The blue polyline indicates the actual values, and the yellow dashed line indicates the predicted values. The figure shows that the model can accurately predict the torsion angle for most time steps, and the predictions are particularly precise in segments with large fluctuations.

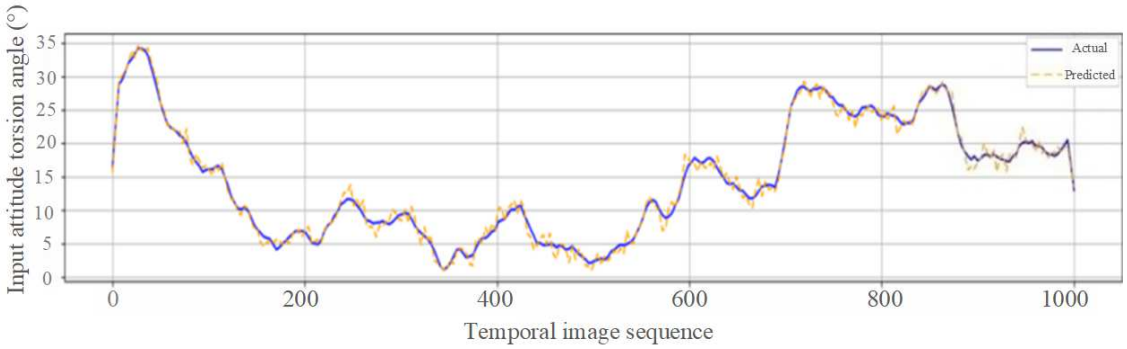


Fig. 9 Comparison between ground truth and predictions

The error curve is shown in Figure 10, where the error is defined as the difference between the ground-truth values and the predicted values. The observed error fluctuations indicate that the model exhibits noticeable deviations at a small number of extreme sample points. Although the overall error remains within a relatively small range, the predictions show slight instability in several highly fluctuating regions. Further optimization and/or collecting additional data are required to improve the model accuracy.

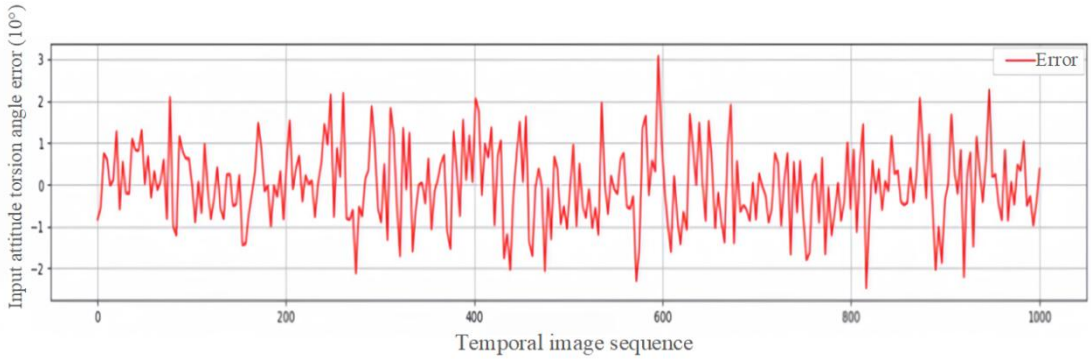


Fig. 10 Line chart of prediction errors

3 Conclusion

This study addresses the problem of belt torsion prediction for pipe belt conveyors in industrial scenarios and proposes a conveyor-belt torsion prediction method based on a spatiotemporal fusion model. The main contributions are summarized as follows:

(1) A spatiotemporal fusion modeling framework and key technical route are proposed for belt torsion prediction. To overcome the limitation of conventional methods that focus primarily on a single temporal sequence while ignoring spatial morphological differences across pipe-forming positions, this work emphasizes synchronous extraction of spatial geometric features and temporal evolution patterns from image sequences. Moreover, transient response, short-term evolution, and long-term trends are integrated into a unified prediction paradigm, providing the foundation for subsequent model design and experimental validation.

(2) A HybridConvLSTM spatiotemporal fusion prediction model is developed and validated, achieving a balance between accuracy and practical real-time performance. The model employs lightweight convolutions to extract local spatial geometric representations and uses ConvLSTM to characterize temporal dynamics, thereby realizing joint spatial–temporal modeling. Comparative experiments and ablation studies on an industrial dataset demonstrate that the proposed method achieves 89.7% accuracy and an F1 score of 83.2% on the test set, significantly outperforming baseline models. The ablation results further confirm the contributions of joint spatiotemporal modeling and lightweight convolution strategies to performance improvement and complexity control.

(3) The experimental pipeline covers the full process of data acquisition, training, deployment, and interpretability analysis, enhancing reproducibility and practical usability. An on-site industrial dataset is constructed, comprising 1000 time-series segments with approximately 3 million frames in total. Stratified sampling and a temporal alignment strategy for dropped frames are adopted to ensure reliable data splitting and evaluation. In addition, the model is optimized on an edge device using TensorRT, improving inference speed from 18 FPS to 28 FPS, which meets real-time industrial early-warning requirements and provides direct evidence for engineering deployment.

CRediT Author Contribution

All authors contributed to the study conception and design. Material preparation, data collection and analysis were performed by Nini Hao, Guoqing Liu, Chaoyue Yu, Xianghui Jia and Yuan Zhang. The first draft of the manuscript was written by Guoqing Liu and all authors commented on previous versions of the manuscript. All authors read and approved the final manuscript.

Funding

This work was supported by the Open Fund of Shandong Province Mine Machinery Engineering Technology Research Center (Grant No. 2022KLMM303), the National Key Research and Development Program of China (Grant No. 2023YFC2907304), and the Taian Science and Technology Innovation Major Project (Grant No. 2022ZDZX010).

Data availability

The data that support the findings of this study are available from the corresponding author upon reasonable request.

Declaration of competing interest

The authors declare that they have no conflict of interest.

References

[1] Wang Y, Du Q G, Han G, et al.environment-protecting con- tinual conveyor -pipe belt convey or [J].Journal of Mechanical Engineering, 2003, (01):149-158.

[2] Chen X B, Zhong Q Y, Liao H, et al.Failure detection technology of inspection robot for tub ular belt conveyor[J].Hoisting and Conveying Machinery,2025,(01):85-90.

[3] Rumin P, Kotowicz J, Hogg D, Zastawna-Rumi n A. Utilization of measurements, machine lear ning, and analytical calculation for preventing b elt flip over on conveyor belts [J]. Measureme nt, 2023, 218: 113157.

[4] dos Santos e Santos L, Ribeiro Filho P R C F, Macêdo E N. Belt rotation in pipe conveyors: Development of an overlap monitoring system using digital twins, industrial Internet of things, and autoregressive language models [J]. Measu rement, 2024, 230: 114546.

[5] Sun F, Hao W, Zou A, Shen Q. A survey on spatio-temporal series prediction with deep learning: taxonomy, applications, and future directions [J]. Neural Computing and Applications, 2024, 36: 9919-9943.

[6] Niazi S, Muqaibel A H, Karimi H R. Multi-scale time series analysis using TT-ConvLSTM technique for bearing remaining useful life prediction [J]. Mechanical Systems and Signal Processing, 2024, 206: 110888.

[7] Qing L ,Zhen P ,Zhaojie L , et al.Structural health monitoring of belt conveyor based on distributed optical fiber acoustic sensor[C]//Guizhou Power Grid Co.,Ltd.(China);Univ. of Electronic Science and Technology of China (China),2022.

[8] Xia Z, Ye F, Dai M, et al. Real-time fault detection and process control based on multi-channel sensor data fusion[J]. The International Journal of Advanced Manufacturing Technology, 2021, 115(3): 795-806.

[9] Sun X, Wang Y, Meng W. Evaluation System of Curved Conveyor Belt Deviation State Based on the ARIMA-LSTM Combined Prediction Model[J]. Machines, 2022, 10, 1042.

[10] Wu X, Wang C, Tian Z, Huang X, Wang Q. Research on belt deviation fault detection technology of belt conveyors based on machine vision [J]. Machines, 2023, 11(12): 1039.

[11] Zhang M, Jiang K, Cao Y, Li M, Wang Q, Li D, Zhang Y. A new paradigm for intelligent status detection of belt conveyors based on deep learning [J]. Measurement, 2023, 213: 112735.

[12] Zhang X, Yang Z, Zhang M, Yu Y, Zhou M, Zhang Y. Quantitative monitoring method for conveyor belt deviation status based on attention guidance [J]. Applied Sciences, 2024, 14(16): 6916.

[13] Hu Xue, Zong Ming. Fault Prediction Method of Belt Conveyor Based on Grey Least Square Support Vector Machine[C]// International Conference on Measuring Technology and Mechatronics Automation. ICMTMA 2021: 55-58.

[14] Qiu M, Li W, Jiang F, et al. Remaining Useful Life Estimation for Rolling Bearing with SIOS-based Indicator and Particle Filtering[J]. IEEE Access, 2018:1-1.

[15] V Molnár, G Fedorko, L Homolka, et al. Utilisation of Measurements to Predict the Relationship between Contact Forces on the Pipe Conveyor Idler Rollers and the Tension Force of the Conveyor Belt[J]. Measurement, 2019, 136.

[16] Chen L, Li X, Du H, Mao F, Zhu H, Xuan J, Zhao Y, Huang Z, Mo K, Zheng Y. An interpretable LAI time series prediction model of subtropic forests using a ConvLSTM coupling spatiotemporal attention mechanism model [J]. Ecological Indicators, 2025, 181: 114437.

[17] Bai S P, Chen G L, Liang K X. Research on common fault monitoring technology of belt conveyor based on BP neural network[J].Opencast Mining Technology,2023,38(03):101-103.

[18] Lin Z, Li M, Zheng Z, et al. Self-attention convlstm for spatiotemporal prediction[C]//Proceedings of the AAAI conference on artificial intelligence. 2020, 34(07): 11531-11538.

[19] Fahad Alharbi,Suhuai Luo,and Guang Yang."TD-CLNet: a time-distributed CNN-LSTM network for fault detection in belt conveyor idlers."Neu

ral Computing and Applications 37.30(2025):doi: 10.1007/S00521-025-11570-2.

[20] Raimondo D M, Limon D, Lazar M, et al. Min-max model predictive control of nonlinear systems: A unifying overview on stability[J]. European Journal of Control, 2009, 15(1): 5-21.

[21] Zhu S T, Liu L T, Niu X H. Applicability of Interpolation Methods for Processing Physics Experiment Data[J].Physical Experiment of College2025,38(06):77-82.

[22] Ouyang P, Kuang X Y, Ye J Z. Path planning based on improved beetle swarm optimization algorithm and cubic spline interpolation[J].Journal of Electronic Measurement and Instrumentation, 2025,39(07):259-268.

[23] Arnab A, Dehghani M, Heigold G, et al. Vivit: A video vision transformer[C]//Proceedings of the IEEE/CVF international conference on computer vision. 2021: 6836-6846.

[24] Chattopadhyay A, Sarkar A, Howlader P, Balasubramanian V N. Grad-CAM++: Improved visual explanations for deep convolutional networks [C]// Proceedings of the IEEE Winter Conference on Applications of Computer Vision (WACV). 2018.

[25] Fayyaz A M, Abdulkadir S J, Talpur N, Al-Selwi S M, Hassan S U, Sumiea E H. Grad-CAM (Gradient-weighted Class Activation Mapping): A systematic literature review [J]. Computers in Biology and Medicine, 2025, 198(Part B): 111200.