ENHANCED YOLOv8-BASED DETECTION OF SURFACE DAMAGES ON CONVEYOR BELTS WITH IMPROVED ACCURACY AND EFFICIENCY

by

Xiuzhuang MEI, Xingtan WANG, Gang XU*, and Ying WANG

School of Mechanical Engineering, Inner Mongolia University of Technology, Hohhot, China

Original scientific paper
https://doi.org/10.2298/TSCI2602919M

Damage to the surface of composite conveyor belts is a prevalent issue that can lead to functional failure. In this study, the YOLOv8 algorithm is enhanced. The convolutional block attention module attention mechanisms and deformable convolutions are introduced, and a small-object detection layer is added. The Focaler--CIoU loss function is also employed. These enhancements are designed to enhance the recognition of damage. In the context of complex operational conditions, the findings indicate a mean recognition accuracy of 93.8% for conveyor belt surface damage, thereby substantiating the efficacy of the enhanced algorithm.

Key words: *conveyor belt, surface damage, YOLOv8, attention mechanism*

Introduction

Belt conveyors are utilized extensively in industries such as coal mining due to their substantial carrying capacity, extensive transportation distances, and high efficiency [1]. Conveyor belts, as core components, are typically composed of composite materials. It has been demonstrated that, over time, material friction, mechanical wear, and the impact of debris can result in surface damage or tearing [2]. This, in turn, can pose a threat to production safety. Consequently, the timely identification of conveyor belt surface damage is imperative for ensuring the safety of equipment operation and facilitating timely repairs.

Recent advancements in deep learning have led to significant progress in the field of conveyor belt surface damage detection. For instance, Zhang *et al.* [3] enhanced YOLOv3 with EfficientNet, thereby elevating the accuracy by 10.4%. Wang *et al.* [4] developed a dual--station model that achieved 98.7% accuracy by expanding the view. Guo *et al.* [5] employed CycleGAN to generate damage images for the purpose of dataset enrichment. Zhang *et al.* [6] proposed a method for enhancing detection speed that is based on Yolov4-tiny. Qing *et al.* [7] incorporated CSPNet in YOLOv4, achieving a 92.5% accuracy rate. Wang *et al.* [8] have made significant advancements in the field of object detection by enhancing the YOLOv7 model for the purpose of detecting longitudinal tears. However, the majority of studies have been conducted in laboratory settings, overlooking the intricate industrial challenges such as coal dust, water stains, and inconsistent lighting. The following paper presents an enhanced YOLOv8-based method for the detection of surface damage on conveyor belts. The model incorporates convolutional block attention module (CBAM) and deformable convolution to fa-

* Corresponding author, e-mail: 1437923111@qq.com

cilitate enhanced feature extraction, incorporates a small object detection layer, and utilizes Focaler-*IoU* to substitute for the original *IoU*, thereby enhancing performance in complex industrial environments, achieving a 24.2% accuracy increase.

Enhanced YOLOv8 detection algorithm

The YOLOv8 is a one-stage object detection approach [9]. This approach transforms the object detection task into a regression problem, thereby enabling a single neural network to process the entire image and generate object location and category details [10]. In this paper, the network structure of YOLOv8 is optimized. The Backbone is augmented with three sets of deformable convolutions (DefConv) and the CBAM. These are placed before and after the C2f modules between convolution layers (Conv). Furthermore, the original 20 × 20 detection layer in the head is substituted with a 160 × 160 detection layer to enhance the ability to detect small objects. The enhanced network configuration is illustrated in fig. 1.

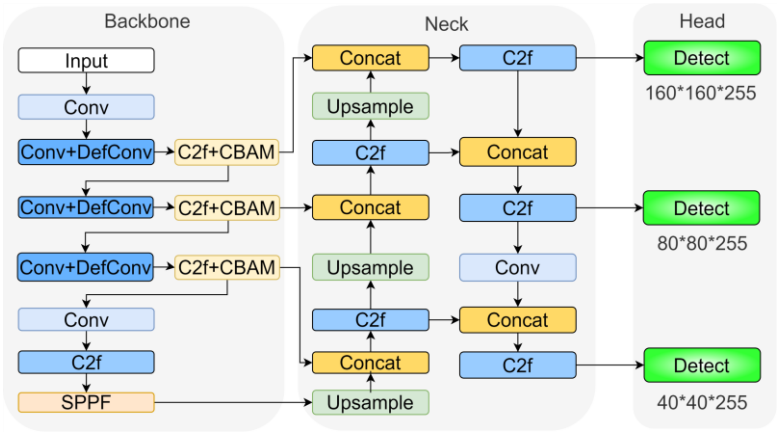

Figure 1. Improved network structure of YOLOv8

Attention mechanism

The attention mechanism enables the network to concentrate on crucial information by imitating the selective attention function of human vision [11]. In this study, a lightweight CBAM is proposed, comprising a channel attention module (CAM) and a spatial attention module (SAM) [12]. The CBAM has been demonstrated to effectively emphasize target features and location information while reducing interference from background elements such as coal ash and water marks. As illustrated in fig. 2, the input feature map undergoes a sequence of convolutional auto-encoder (CAE) and shared auto-encoder (SAM) processing. Following the integration of these processes, an output feature map is generated.

The CAM is a statistical technique that calculates the relationship among features. Initially, the global maximum pooling and global average pooling operations are performed on each channel of the input feature map. This results in two channel maps of size 1 × 1 × C. The maps are processed through a shared fully-connected neural network (Shared MLP), which produces two 1-D vectors. Subsequently, these vectors are aggregated and processed by an activation function to yield the final channel attention vector. The configuration of the CAM module is illustrated in fig. 3.

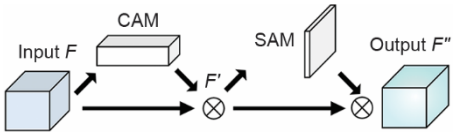

Figure 2. Structure of the CBAM module

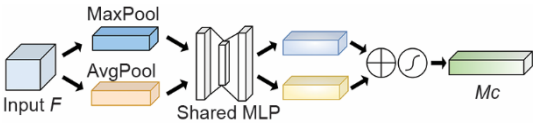

Figure 3. Structure of the CAM module

The CAM calculation formula is:

$$Mc = \sigma\{MLP[AvgPool(F)] + MLP[MaxPool(F)]\}$$
$$F' = Mc \otimes F \tag{1}$$

where $AvgPool(\cdot)$ and $MaxPool(\cdot)$ represent global average and max pooling, $MLP(\cdot)$ denotes the fully-connected layer, and $\sigma(\cdot)$ signifies the Sigmoid activation operation.

The SAM algorithm is designed to generate an attention feature map, a process that involves the analysis of spatial relationships within channels. Firstly, the feature map processed by the CAE undergoes gradient-based methods, including gradient ascent and gradient descent, along the channel direction. The results are then processed using a 7 × 7 convolution kernel to yield the spatial attention vector. The vector values are subsequently normalized to the 0-1 range through the Sigmoid activation function to form the spatial attention vector. The configuration of the SAM module is illustrated in fig. 4.

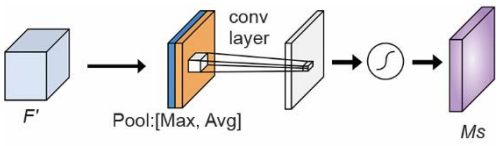

Figure 4. Structure of the SAM module

The SAM calculation formula is:

$$Ms = \sigma\{f^{7\times7}\{[AvgPool(F'); MaxPool(F')]\}\}$$
$$F'' = Ms \otimes F' \tag{2}$$

In the Backbone, CBAM is added after the C2f modules between Conv modules, facilitating the effective transmission of enhanced target features to the Neck section.

Deformable convolution

Standard convolutions are characterized by fixed shapes and receptive fields, which impose limitations on their capacity to extract features from targets exhibiting large width-height ratios. Deformable convolution (DefConv) is a technique that has been developed for the purpose of enhancing the feature extraction capacity of convolutions. In this method, each element of the convolution kernel is assigned a corresponding directional parameter, thereby adapting to the target's shape and size. This augmentation of the receptive field is accompanied by an accommodation of geometric feature transformations. The computational process of deformable convolution is illustrated in fig. 5.

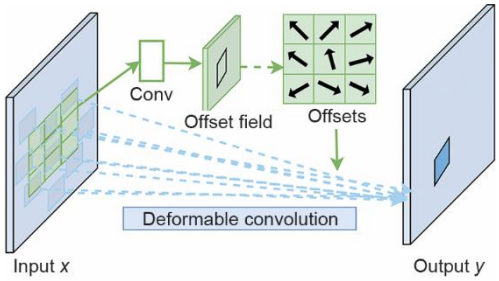

Figure 5. Illustration of deformable convolution

The formula for deformable convolution is:

$$y(p_0) = \sum_{p_n \in R} w(p_n) x(p_0 + p_n + \Delta p_n) \tag{3}$$

where x is the input feature map, y – the output feature map, p_0 – the current convolution calculation position, p_n – the n^{th} corresponding position, and Δp_n – the positional offset. The specific calculation steps include:

– Computing the offset via convolution branches on the input image.
– Determining the positions of network-adaptive sampling points based on the offset.
– Obtaining feature values of these sampling points using bilinear interpolation.
– Performing a weighted summation to produce the output feature map.

The implementation of this mechanism prior to certain C2f modules in the Backbone has been demonstrated to enhance the model's capacity to discern intricate details. Furthermore, it has been demonstrated that this approach enhances the model's generalization capabilities with respect to spatial adaptability and geometric transformations.

Focaler-CIoU loss function

The CIoU loss in YOLOv8

The YOLOv8 uses the complete intersection over union loss (CIoU) as the regression loss function for bounding boxes. The calculation formula is:

$$L_{CIoU} = 1 - IoU + \frac{\rho^2(B_{\text{gt}}, B_{\text{pred}})}{C^2} + \alpha v \tag{4}$$

IoU, v, and α are defined:

$$IoU = \frac{B_{\text{gt}} \cap B_{\text{pred}}}{B_{\text{gt}} \cup B_{\text{pred}}}, \quad v = \frac{4}{\pi^2}\left(\arctan\frac{w_{\text{gt}}}{h_{\text{gt}}} - \arctan\frac{w_{\text{pred}}}{h_{\text{pred}}}\right)^2, \quad \alpha = \frac{v}{(1 - IoU) + v} \tag{5}$$

where B_{gt} is the ground truth box, B_{pred} – the predicted box, $\rho(B_{\text{gt}}, B_{\text{pred}})$ – the distance between the centers of the predicted and ground truth boxes, C – the diagonal distance of the smallest enclosing rectangle of the predicted and ground truth boxes, and v – the consistency in aspect ratios of the predicted and ground truth boxes. The w_{gt}, h_{gt}, w_{pred}, and h_{pred} represent the width and height of the ground truth and predicted boxes, while α is the weight value. When remains constant, a larger *IoU* implies a greater α, indicating that a high *IoU* places more emphasis on the aspect ratio consistency between predicted and ground truth boxes.

Improved Focaler-CIoU loss function

To better focus on small objects and facilitate learning, the Focaler-*IoU* method is integrated into the loss function. Focaler-*IoU* reconstructs *IoU* using a linear interval mapping method, changing the regression ratio of detection boxes. The reconstructed *IoU* formula is:

$$IoU^{\text{Focaler}} = \begin{cases} 0, & IoU < d \\ \dfrac{IoU - d}{u - d}, & d \le IoU \le u \\ 1, & IoU > u \end{cases} \tag{6}$$

where *u* and *d* denote the upper and lower bounds of *IoU*. Based on eq. (6), the Focaler-*IoU* mapping curve can be plotted, as shown in fig. 6.

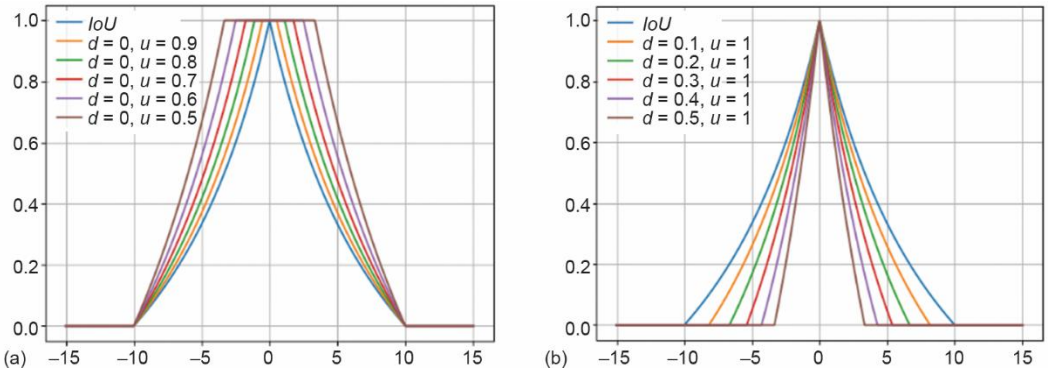

Figure 6. Focaler-*IoU* mapping curve; (a) *d* = 0 and (b) *u* = 1

Figure 6(a) shows the Focaler-*IoU* mapping curve obtained by reducing u with a fixed *d*. Conversely, fig. 6(b) presents the curve resulting from increasing *d* with a fixed *u*. The figures indicate that when *d* is fixed at 0, a smaller u correlates to a smaller overlap area under the same *IoU* conditions. When *u* is set to 1, despite equal overlap areas, a larger *d* leads to a smaller *IoU* value. To increase the loss proportion for small objects, *u* is set to 0.9 and *d* to 0.

Replacing the original loss function's *IoU* with IoU^{Focaler} with results in the final localization loss function:

$$L_{\text{Focaler}-CIoU} = 1 - IoU^{\text{Focaler}} + \frac{\rho^2(B_{\text{gt}}, B_{\text{pred}})}{C^2} + \alpha v \tag{7}$$

Using Focaler-*CIoU* loss as the model loss function effectively increases the loss proportion for small objects and accelerates the model convergence speed.

Addition of small object detection layer

The detection head of YOLOv8 focuses on object localization, outputting the center coordinates and size information of objects. The model has three output sizes, namely 20 × 20, 40 × 40, and 80 × 80, to provide suitable resolution for detecting large, medium, and small objects.

When the pixel area of an object in the image is 50 × 50, it is usually considered a small object. Under such circumstances, YOLOv8 detection performance deteriorates. To enhance the network small-object recognition ability, the 20 × 20 detection layer in the detection head is replaced with a 160 × 160 detection layer. This adjustment aims to preserve more small-object feature information and improve the model ability to accurately capture and identify small objects, thereby enhancing the overall detection performance.

Experiments and results analysis

Experimental environment and parameter settings

The experiments were conducted on a Windows 10 operating system. The computational power was provided by a Tesla P100 GPU and an Intel Xeon E5-2678 CPU. The

PyTorch deep learning framework was adopted for model implementation. The image resolution was set to 640 × 640. A batch size of 16 was used, along with a learning rate of 0.01. The Adam optimizer was selected, and a weight decay coefficient of 0.00125 was applied. The number of iterations was 1000, and the *IoU* threshold was set at 0.7.

Experimental dataset

A dataset consisting of 3000 images showing conveyor belt surface damage was used for training purposes. This dataset was randomly partitioned into training and testing subsets at an 8:2 ratio.

Ablation study

An ablation study was designed to examine the influence of the various improvements on the model performance. The results of this experimental study are presented in tab. 1.

Table 1. Ablation study results

Number	Attention mechanism	Deformable convolution	Focaler-*CIoU*	Small object	*P*	*R*	FLOP	fps
0					0.696	0.678	**8.9**	**112.9**
1	1				0.785	0.695	8.9	109.9
2	1	1			0.865	0.815	9.7	84.9
3	1	1	1		0.913	**0.901**	9.7	88.6
4	1	1	1	1	**0.938**	0.824	15.8	47.2

In tab. 1, *P* is the precision, *R* – the recall, FLOP – the computational load, and fps – the frames per second. From the results, it can be seen that deformable convolutions and attention mechanisms are capable of improving both *P* and *R*. However, this improvement in performance comes at the cost of an increase in computational complexity, which causes a decrease in fps. Nevertheless, the final model exhibits a significant 24.2% improvement in precision, clearly demonstrating its effective detection capabilities.

Comparative experiment

A comparative analysis was carried out between the improved algorithm and the currently prevalent mainstream algorithms. The results of this comparison are presented in tab. 2.

Table 2. Comparative experiment results

Number	Model	*P*	*R*	FLOP	fps
0	YOLOv5	0.626	0.589	**7.6**	61.7
1	YOLOv7	0.739	0.624	8.8	70.8
2	YOLOv8	0.696	0.678	8.9	**112.9**
3	SSD	0.882	**0.844**	18.9	46.3
4	Faster R-CNN	0.899	0.837	26.4	25.8
5	Ours	**0.938**	0.824	15.8	47.2

According to the data presented in tab. 2, in contrast to Faster R-CNN, the improved algorithm led to a 21.4% increase in fps. When compared with *SSD*, there was a 5.6% enhancement in *P*. Among similar series of algorithms, compared to YOLOv5, YOLOv7, and YOLOv8, respectively, *P* of the improved algorithm increased by 31.2%, 19.9%, and 24.2%. These outcomes suggest that the proposed improved algorithm demonstrates enhanced detection performance compared to other algorithms.

Detection effectiveness analysis

A comparative analysis was performed between the proposed algorithm and YOLOv8, and examples of conveyor belt surface damage detection are illustrated in figs. 7-9.

In the small-target detection scenario shown in fig. 7(a), YOLOv8 exhibits certain limitations when it comes to detecting small targets. It can be observed from fig. 7(b) that YOLOv8 has difficulty in capturing the fine-grained damage features. On the contrary, by incorporating a small object detection layer, the proposed algorithm remarkably enhances the detection accuracy for small-sized damage targets. As is evident in fig. 7(c), the proposed algorithm is capable of successfully detecting the minute damage on the conveyor belt surface.

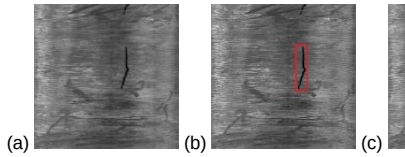

Figure 7. Conveyor belt surface damage detection Example 1; (a) original, (b) YOLOv8 result, and (c) ours result

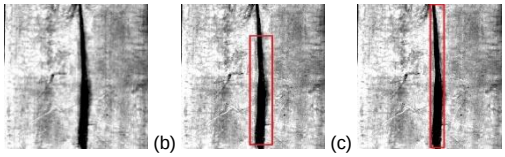

Figure 8. Conveyor belt surface damage detection Example 2; (a) original, (b) YOLOv8 result, and (c) ours result

In the case of the damage target with an overly large aspect ratio presented in fig. 8(a), YOLOv8 is unable to extract complete features because of its restricted receptive field. This leads to incomplete recognition, as can be witnessed in fig. 8(b). However, by integrating deformable convolution, the proposed algorithm enables the convolution elements to adaptively adjust their positions. This effectively expands the receptive field and accomplishes accurate recognition of targets with complex shapes. Figure 8(c) clearly demonstrates the algorithm success in fully recognizing damage targets that have larger longitudinal proportions.

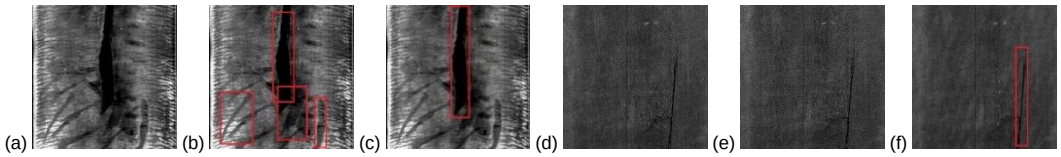

Figure 9. Conveyor belt surface damage detection Example 3; (a) original, (b) YOLOv8 result, (c) ours result; (d) original, (e) YOLOv8 result, and (f) ours result

In scenarios where coal dust and water stains are present, as illustrated in figs. 9(a) and 9(d), YOLOv8 demonstrates a tendency to produce false positives and missed detections, as evidenced in figs. 9(b) and 9(e). This finding suggests that these factors have a substantial impact on the performance of the detection system. The proposed algorithm integrates an attention mechanism that focuses on the features of damage targets during the Backbone stage and further extracts features during the Neck stage. This property enables the system to accu-

rately distinguish between damage targets and the background. Notably, the proposed algorithm exhibits a remarkable capacity to maintain high-precision recognition, successfully avoiding both false positives and missed detections, as evidenced by the experimental results depicted in figs. 9(c) and 9(f).

The aforementioned examples demonstrate that the enhanced network exhibits exceptional recognition accuracy and robust anti-interference capabilities, thereby ensuring its reliable operation under complex working conditions.

Conclusions

This study proposes an enhanced algorithm, termed YOLOv8, which was developed for the purpose of detecting surface damage on conveyor belts. The algorithm under review includes several enhancements of particular significance. The CBAM attention mechanism has been integrated into the YOLOv8 backbone. This integration effectively mitigates the interference from background elements, including coal dust and water stains, thereby enabling accurate differentiation between damage targets and background information. Deformable convolution has been introduced, thus bolstering the model capacity to handle spatial deformations. Consequently, the model is capable of accurately and comprehensively extracting features of damage targets with large aspect ratios. The loss function has been refined to Focaler-*CIoU*, which optimizes sample balance and expedites the convergence speed. A small object detection layer has been added, endowing the model with the ability to effectively detect smaller damage targets. Collectively, these improvements have led to a notable enhancement in the detection accuracy, reaching 93.8% under complex working conditions. This ensures the model's applicability within a wide range of challenging environments.

Subsequent research endeavors will concentrate on AI-driven control loops [13] and sophisticated deep learning methodologies [14-16] to facilitate real-time adaptive modeling and predictive maintenance capabilities, thereby reinforcing the synergy between computer vision and industrial automation systems. Additionally, surface damage characterization [17] can be quantitatively analyzed through a two-scale fractal dimension modeling approach [18-21].

Acknowledgment

This research was supported by the Basic Scientific Research Business Fund Projects of Universities directly under the Inner Mongolia Autonomous Region. Specific projects include the Inner Mongolia University of Technology Scientific Research Startup Fund Project (DC2200000929) and the Inner Mongolia University of Technology's Discipline Team for Reliability and Optimal Design of Complex Equipment (JY20230094). Special thanks are extended to the reviewers for their valuable feedback on this paper.

References

[1] Mei, L. S., *et al.*, An On-Line Detection Method for Conveyor Belt Deviation Faults, *Thermal Science*, *27* (2023), 3A, pp. 2099-2107
[2] Mei, M. Y., *et al.*, On-Line Intelligent Evaluation of the Fatigue State of a Composite Conveyor Belt, *Thermal Science*, *25* (2021), 3B, pp. 2191-2198
[3] Zhang, M., *et al.*, Deep Learning-Based Damage Detection of Mining Conveyor Belt, *Measurement*, *175* (2021), Apr., pp. 109-130
[4] Wang, G. L., *et al.*, Multi-View Detection Method for Longitudinal Tear of Conveyor Belt Based on Improved YOLOv3, *Journal of Hefei University of Technology (Natural Science)*, *46* (2023), 1, pp. 28-35+80
[5] Guo, X., *et al.*, Damage Detection for Conveyor Belt Surface Based on Conditional Cycle Generative Adversarial Network, *Sensors (Basel)*, *22* (2022), 9, pp. 1-17

[6] Zhang, *et al.*, Damage Detection Method for Mine Conveyor Belt Based on Deep Learning, *Industry and Mine Automation*, *47* (2021), 6, pp. 51-56
[7] Qing, L., *et al.*, Longitudinal Tear Detection Based on Improved YOLOv4 for Conveyor Belt, *Computer System Applications*, *32* (2023), 3, pp. 186-194
[8] Wang, Y. M., *et al.*, Longitudinal Tear Detection of Conveyor Belt Based on Improved YOLOv7, *Ieee Access*, *12* (2024), Feb., pp. 24453-24464
[9] Li, *et al.*, YOLOv8 Ship Target Detection Algorithm Based on CCA and Transformer, *Control Engineering of China*, *31* (2024), 5, pp. 1-12
[10] Yang, L. J., *et al.*, Wear State Detection of Conveyor Belt in Underground Mine Based on Retinex-YOLOv8-EfficientNet-NAM, *Ieee Access*, *12* (2024), Feb., pp. 25309-25324
[11] Wang, *et al.*, Research on Pedestrian Detection Algorithm Based on Mixed Attention Mechanism and C2f, *Software Guide*, *23* (2024), 1, pp. 135-142
[12] Woo, S., *et al.* CBAM: Convolutional Block Attention Module, *Proceedings*, 15th European Conference on Computer Vision (ECCV), Munich, Germany, Vol. 11211, 2018, pp. 3-19
[13] He, J. H. Transforming Frontiers: The Next Decade of Differential Equations and Control Processes. *Advances in Differential Equations and Control Processes*, *32* (2025), 1, 2589
[14] Nosonovsky, M., Aglikov, A. S., Triboinformatics: Machine Learning Methods for Frictional Instabilities, *Facta Universitatis, Series Mechanical Engineering 22* (2024), 3, pp. 423-433
[15] Cao, Y. J., *et al.* Study of Friction Compensation Model for Mobile Robot's Joints, *Facta Universitatis, Series Mechanical Engineering*, *22* (2024), 4, pp. 721-740
[16] Karim, F. K., *et al.*, Innovative Mathematical Modelling Approaches to Diagnose Chronic Neurological Disorders with Deep Learning, *Thermal Science*, *28* (2024), 6B, pp. 5217-5229
[17] Chen, Z. P., *et al.*, Multi-Role Collaborative Framework for Structural Damage Identification Considering Measurement Noise Effect, *Measurement*, *250* (2025), 117106
[18] Guan, Y. Z., *et al.*, Variational Formulations for a Coupled Fractal-Fractional KdV System, *Fractals*, *32* (2024), 2450054
[19] Li, X. X., *et al.*, Elucidating the Fractal Nature of the Porosity of Nanofiber Members in the Electrospinning Process, *Fractals*, *32* (2024), 2450109
[20] Liu, Y. P., *et al.*, Leveraging Lotus Seeds' Distribution Patterns for Fractal Super-Rope Optimization, *Fractals*, *33* (2025), 2450143
[21] Liu, Y. P., *et al.*, Clover-Inspired Fractal Architectures: Innovations in Flexible Folding Skins for Sustainable Buildings, *Fractals*, *33* (2025), 2550041

Paper submitted: October 26, 2024
Paper revised: March 28, 2025
Paper accepted: April 1, 2025