mathematics

MDPI

Article

Adaptive Variational Modal Decomposition–Dual Attention Mechanism Parallel Residual Network: A Tool Lifetime Prediction Method Based on Adaptive Noise Reduction

Jing Kang [1], Taiyong Wang [1,2], Yi Li [1], Ye Wei [1], Yaomin Zhang [1] and Ying Tian [1,*]

[1] School of Mechanical Engineering, Tianjin University, Tianjin 300354, China; 1021201038@tju.edu.cn (J.K.); tywang@tju.edu.cn (T.W.); liyi0902@foxmail.com (Y.L.); yewei0116@tju.edu.cn (Y.W.); zhangyaom@tju.edu.cn (Y.Z.)
[2] Renai College, Tianjin University, Tianjin 301636, China
* Correspondence: tianying@tju.edu.cn

Abstract: This paper addresses the issue of noise interference and variable working conditions in the production and machining environment, which lead to weak tool life features and reduced prediction accuracy. A tool lifetime prediction method based on AVMD-DAMResNet is proposed. The method first adapts the parameters of the variational modal noise reduction algorithm using an improved sparrow optimization algorithm, and then reconstructs the original vibration signal with noise reduction. Second, the residual module of the deep residual network is enhanced using a two-dimensional attention mechanism. A parallel residual network tool prediction model (DAMResNet) was constructed to optimize the model's weight allocation to different features, achieving multi-channel and multi-dimensional feature fusion. Finally, the noise-reduced signal was input into the DAMResNet model to accurately predict tool lifetime. The experimental results show that, compared with the original ResNet model, the proposed AVMD-DAMResNet model improves the coefficient of determination (R^2) by 5.8%, reduces the root mean square error (RMSE) by 31.2%, and decreases the mean absolute percentage error (MAPE) by 31.4%. These results demonstrate that the AVMD-DAMResNet-based tool lifetime prediction method effectively reduces noise and achieves high prediction accuracy.

Keywords: tool lifetime prediction; signal denoising; variational modal decomposition; residual neural network

MSC: 68T07; 68T09

check for updates

Academic Editors: Qilin Li and Senjian An

Received: 30 November 2024
Revised: 18 December 2024
Accepted: 23 December 2024
Published: 25 December 2024

Citation: Kang, J.; Wang, T.; Li, Y.; Wei, Y.; Zhang, Y.; Tian, Y. Adaptive Variational Modal Decomposition–Dual Attention Mechanism Parallel Residual Network: A Tool Lifetime Prediction Method Based on Adaptive Noise Reduction. *Mathematics* **2025**, *13*, 25. https://doi.org/10.3390/math13010025

Copyright: © 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by/4.0/).

1. Introduction

In actual production and processing environments, the complexity of working conditions, including significant environmental noise, leads to interference from external vibration sources during signal acquisition. As a result, the acquired signal contains not only the characteristic frequencies of the measured workpiece but also substantial interference, which diminishes the significance of tool lifetime features and reduces prediction accuracy. Therefore, deep learning-based tool lifetime prediction methods typically require feature extraction or signal noise reduction techniques applied to the raw signal [1]. Traditional methods such as time–frequency domain feature extraction, wavelet noise reduction, and empirical modal decomposition rely heavily on expert knowledge for decomposition, which limits the prediction accuracy of deep learning models. Therefore, designing a tool

life prediction model that is robust to noise interference and capable of accurate predictions is crucial.

Time domain feature extraction involves analyzing the waveform, amplitude, and other characteristic parameters of the signal to identify critical indicators reflecting changes in tool lifetime. Commonly used time domain parameters include mean value, kurtosis, and variance. Jemielniak et al. [2] proposed the signal features (SFs) are extracted from time domain signals, frequency domain transforms, and wavelet coefficients (the time–frequency domain). Their relevance for tool wear monitoring is automatically evaluated using the coefficient of determination, low-pass filtering, and repeatability analysis. Wang et al. [3] used wavelet packet decomposition to process cutting force signals collected during milling, extracting energy values of different signal components to construct a model for predicting tool wear. Guo et al. [4] proposed a novel similarity measure combining Euclidean distance and cosine similarity, called the Representative IMF Selection Index (RISI) method, to identify IMFs with significant degradation features. An LSTM model was then trained for each selected IMF and its residuals. Sayyad et al. [5] considers time–frequency domain (TFD) features, such as the short-time Fourier transform (STFT) and different wavelet transforms (WTs), as well as deep learning (DL) models, such as the long short-term memory (LSTM), different variants of the LSTN, convolutional neural networks (CNNs), and hybrid models (combinations of CCNs and variants of LSTMs), for the prediction of the remaining lifetime. Tool vibration signal decomposition and reconstruction typically involve methods such as wavelet transform, variational modal decomposition (VMD), and empirical modal decomposition (EMD). These methods decompose the signal, select strongly correlated components for reconstruction, effectively preserve the signal's original features, and facilitate deep learning models in extracting deep features from the signal.

In tool wear and life monitoring, signal decomposition and reconstruction methods, such as wavelet transform [6] and empirical variational modal decomposition [7], are commonly used to extract strongly correlated signals for reconstruction. This approach preserves the original features of the signals, facilitating the extraction of deep features by deep learning models. Liu et al. [8] applied wavelet transform for noise reduction for tool vibration signals and achieved tool lifetime prediction using an enhanced long short-term memory network with an attention mechanism. Benkedjouh et al. [9] used a wavelet thresholding method for noise reduction and a bi-directional long- and short-term memory network for tool wear state monitoring. Variational mode decomposition (VMD), proposed by Dragomiretskiy et al. in 2014, is a signal analysis method that offers significant advantages in analyzing and processing nonlinear and non-smooth signals [10,11]. VMD overcomes the limitations of wavelet-based decomposition, such as lack of adaptivity, as well as the endpoint effect and modal aliasing associated with empirical mode decomposition (EMD). It provides a solid mathematical foundation, enabling effective noise reduction, feature separation, and the decomposition of complex signals with limited bandwidth and non-overlapping frequency bands. VMD also demonstrates excellent noise reduction for signals containing strong noise. It demonstrates excellent noise reduction. Liu et al. [12] proposed a novel prediction method for machining tools under dynamic operating conditions with varying spindle loads. This method enhances variational mode decomposition to extract components with strong trend-following and noise robustness for feature extraction. Sun et al. [13] combined variational modal decomposition (VMD) with Pearson correlation thresholding (PCT) to adaptively filter machining signals and constructed a stacked bi-directional long short-term memory (AT-SBiLSTM) model with an attention mechanism. This model accurately predicts long-term wear in the milling process using limited monitoring data from a single tool. The above methods effectively preserve the original features of the data, achieve high data utilization, and produce processed data

suitable for training deep learning models. However, some parameters still require manual selection, and the level of automation needs to be improved.

In recent years, deep learning models have made significant advances in tool wear condition monitoring and lifetime prediction, owing to their exceptional ability to adaptively extract deep features [14,15]. Deep learning technology enables the rapid and accurate extraction of correlations between different features and tool lifetime from noise-contaminated machining signals. It also improves the learning weights of strongly correlated features, providing strong support for effective tool condition monitoring and accurate life prediction. Qiao et al. [16] collected tool vibration signals and wear data to implement real-time tool wear prediction using a distributed CNN network. Zhou et al. [17] proposed an improved convolutional neural network with multi-scale edge labeling to enhance tool lifetime prediction accuracy in small sample scenarios. Kumar et al. [18] optimized the tool life prediction model by tuning neural network hyperparameters, resulting in improved model performance. Duan et al. [19] uses short-time Fourier transform to process the signals and input them into a residual neural network for tool lifetime monitoring. The deep learning-based tool lifetime prediction method enhances feature extraction compared to traditional machine learning approaches, improving prediction accuracy and performing excellently under simple conditions, such as single operating conditions and single vibration signals. However, key challenges remain under complex conditions, such as noise interference and variable operating conditions.

In summary, when processing highly sampled and noisy tool signals, eigenvalue extraction methods are prone to noise interference, and the lack of adaptivity in eigenvalue selection hampers model stability. Signal reconstruction methods, such as wavelet analysis, are also constrained by algorithm hyperparameters, such as the basis function, and their effectiveness is influenced by subjective factors, limiting their full potential. As machine tools operate under complex conditions, such as variable speed and load, there is still potential for improving the versatility of deep learning models in handling a single operating condition. Therefore, studying effective tool lifetime prediction methods under noisy and variable conditions is of great practical significance for achieving accurate tool lifetime prediction.

2. Data Preprocessing Methods

2.1. Variational Modal Decomposition

Compared to traditional signal processing methods, such as short-time Fourier transform, wavelet transform, and S transform, the variational modal decomposition (VMD) method is adaptive and fully non-recursive. Unlike empirical modal decomposition (EMD), VMD can autonomously determine the number of intrinsic modal components, and the number of components can be adjusted based on the specific operating conditions. The algorithm constructs and solves a variational decomposition problem to achieve this. In the iterative solving process, the algorithm adaptively selects the optimal center frequency and corresponding bandwidth for each component, effectively separating data features, dividing the signal's time–frequency band, and extracting meaningful decomposition components. This provides improved decomposition accuracy and noise reduction for complex, noise-contaminated data under actual processing conditions. The basic principle of variational modal decomposition is to separate signal features by constructing a frequency domain variational optimization problem to find the optimal solution.

Variational modal decomposition consists of two stages: constructing the constrained variational model and solving it. First, the constrained variational model is constructed using the constraints of the IMF components. The specific procedure is as follows:

Let the original signal be denoted as $x(t)$. First, determine the number of decomposition layers K, which defines the number of IMF components. The k-th IMF component is denoted by $u_k(t)$. To obtain the one-sided spectrum, the Hilbert transform is applied to the signal, resulting in the resolved signal:

$$(\delta(t) + j/\pi t) * u_k(t) \tag{1}$$

where $k \in \{1, 2, \ldots, K\}$, $\delta(t)$ is the delta function and $*$ is the convolution operator.

The center frequency ω_k, corresponding to the resolved signal for each IMF component, is then estimated, and the spectrum is converted to the fundamental band by multiplying the spectral expression of the resolved signal by the exponential form of ω_k. The spectrum is then converted to the fundamental band:

$$[(\delta(t) + j/\pi t) * u_k(t)]e^{-j\omega_k t} \tag{2}$$

Finally, the complete constrained variational equation is constructed, which contains the number of model decomposition layers K, the IMF bandwidth and center frequency constraints, and the error between the sum of each signal component and the original signal to satisfy the accuracy requirement:

$$\begin{aligned} &\min_{\{u_k\},\{\omega_k\}} \left\{ \sum_k \| \, \partial_t [(\delta(t) + j/\pi t) * u_k(t)] e^{-j\omega_k t} \, \|_2^2 \right\} \\ &\text{s.t.} \sum_{k=1}^{k} u_k = f \end{aligned} \tag{3}$$

where k is the number of layers chosen for the modal decomposition, u_k represents the K modal components of the decomposition, and ω_k represents the center frequency of each modal component.

The second part is to solve the constrained variational model, using the Lagrange multiplier operator λ and the quadratic penalty factor α; the above equation is transformed into an unconstrained variational problem and the augmented Lagrange expression is obtained:

$$\begin{aligned} L(\{u_k\}, \{\omega_k\}, \lambda) = \alpha \sum_k \left\| \partial_t [(\delta(t) + j/\pi t) * u_k(t)] e^{-j\omega^k t} \right\|_2^2 + \\ \left\| f(t) - \sum_k u_k(t) \right\|_2^2 + \left\langle \lambda(t), f(t) - \sum_k u_k(t) \right\rangle \end{aligned} \tag{4}$$

Finally, the expression is solved using the alternating direction method, and the optimal solution of the model is obtained by iterating according to the accuracy requirements to obtain the K modal component functions.

Although VMD offers better denoising performance than traditional algorithms, it has certain limitations. The variational modal decomposition requires manually selecting two parameters: the penalty factor and the number of decomposition layers. If these parameters are too small or too large, they can significantly affect the decomposition results, reducing the adaptivity and feature significance of the VMD method.

2.2. Adaptive Variational Modal Decomposition (AVMD) Algorithm

The sparrow search algorithm (SSA) is a swarm optimization algorithm proposed in 2020, inspired by the foraging and anti-predation behaviors of sparrows [20,21]. The algorithm maps the natural environment to the solution space and individual sparrows to datasets within the algorithm. It achieves automatic location optimization by simulating the foraging behavior through an adaptivity function, guiding the search toward optimal solutions. In the algorithm, individuals serve three roles: the "discoverer" leads the

population in a global search, the "follower" trails the discoverer and performs local optimization, and the "alert" role detects danger, halting foraging and adjusting the search direction. The algorithm begins by randomly initializing a set of individuals with position and velocity. The velocity represents the direction and distance of movement in the next iteration, while the position corresponds to a solution, with the dimensionality determined by the problem.

Using $x_{i,j}^t$ to represent the i-th sparrow, provides the value of the j-th coordinate at iteration number t. If we let the dimension of the solution be d, the finder update formula is as follows:

$$x_{i,j}^{t+1} = \begin{cases} x_{i,j}^t \exp\left(\frac{-i}{\alpha iter_{\max}}\right) & R_2 < ST \\ x_{i,j}^t + qL & R_2 \geq ST \end{cases} \tag{5}$$

where q represents a random number obeying a standard normal distribution, α is a random number of (0, 1], $iter_{\max}$ is the maximum number of iterations, and L is an $i \times d$ matrix with all values equaling one. R_2 and ST represent the alarm values and safety thresholds in the algorithm. This part of the structure represents the global search phase of the algorithm, where individuals will perform a fast global optimization under the control of the safety threshold.

Afterwards, based on the fitness parameter, a proportion of the individuals are set as followers, and the position of the followers is updated with the formula:

$$x_{i,j}^{t+1} = \begin{cases} q \exp\left(\frac{x_{worst}^t - x_{i,j}^t}{i^2}\right) & i > \frac{n}{2} \\ x_P^{t+1} + \left| x_{i,j}^t - x_P^{t+1} \right| \mathbf{A} L & i \leq \frac{n}{2} \end{cases} \tag{6}$$

where x_P^{t+1} represents the best position of the current discoverer, x_{worst}^t represents the worst position, $\mathbf{A}$ is a $1 \times d$ matrix with elements randomized to positive and negative 1, and $\mathbf{A}^+ = \mathbf{A}^{\mathrm{T}}\left(\mathbf{A}\mathbf{A}^{\mathrm{T}}\right)^{-1}$. This part represents the local search process of the algorithm.

The algorithm performs random vigilance behavior after completing a finder and follower optimization search, generally 10% of the total number is selected to generate the vigilance individual, and the vigilance individual's position update formula is as follows:

$$x_{i,j}^{t+1} = \begin{cases} x_{best}^t + \beta_1 \left(x_{i,j}^t - x_{best}^t \right) & f_i \neq f_g \\ x_{i,j}^t + \beta_2 \left(\frac{x_{i,j}^t - x_{worst}^t}{f_i - f_w + \varepsilon} \right) & f_i = f_g \end{cases} \tag{7}$$

where x_{best}^t represents the current optimal position, and β_1 and β_2 are the random numbers used to control the step size and direction, respectively, and obey the standard normal distribution. f_i is the fitness value of the i-th individual, f_g and f_w represent the current global optimal fitness and global worst fitness, respectively, and ε is a small constant to prevent the denominator of the equation from being zero.

The sparrow search algorithm initializes the population randomly at the start, which often leads to an uneven distribution of individuals, thereby reducing the algorithm's global search capability. To improve the diversity and uniformity of the initialized population, chaotic mapping is introduced to optimize the initialization process. In this section, the algorithm is enhanced using the Cubic-Tent (CT) initialization method, as shown in the following formula:

$$x_{n+1} = \begin{cases} \beta x_n \left(1 - x_n^2\right) + \frac{(4-\beta) x_n}{2} & x_n < 0.5 \\ \beta x_n \left(1 - x_n^2\right) + \frac{(4-\beta)(1-x_n)}{2} & x_n \geq 0.5 \end{cases} \tag{8}$$

where x_n is the solution at the current position and x_{n+1} is the solution at the next step. β is a control parameter that regulates the extent of the search.

This section proposes an adaptive variational modal decomposition (AVMD) algorithm that optimizes VMD using the improved sparrow search algorithm. The algorithm adapts the signal envelope entropy and automatically selects the optimal penalty factor α and the number of decomposition layers K, thereby improving the decomposition performance of VMD.

The signal noise reduction process based on the AVMD algorithm is shown in Figure 1. The steps are as follows:

(1) Input the original signal into the algorithm.
(2) Initialize the algorithm using CT mapping and assign random positions to the individuals.
(3) Sequentially update the positions of the discoverer, follower, and alert.
(4) Calculate the fitness value based on individual positions. If the accuracy is met or the maximum number of iterations is reached, the algorithm stops and outputs the global optimal solution. Otherwise, repeat the previous steps until the requirements are satisfied.
(5) Output the signal decomposed on the basis of the optimal solution.
(6) Remove weakly correlated components using correlation analysis.
(7) Output the reconstructed signal.

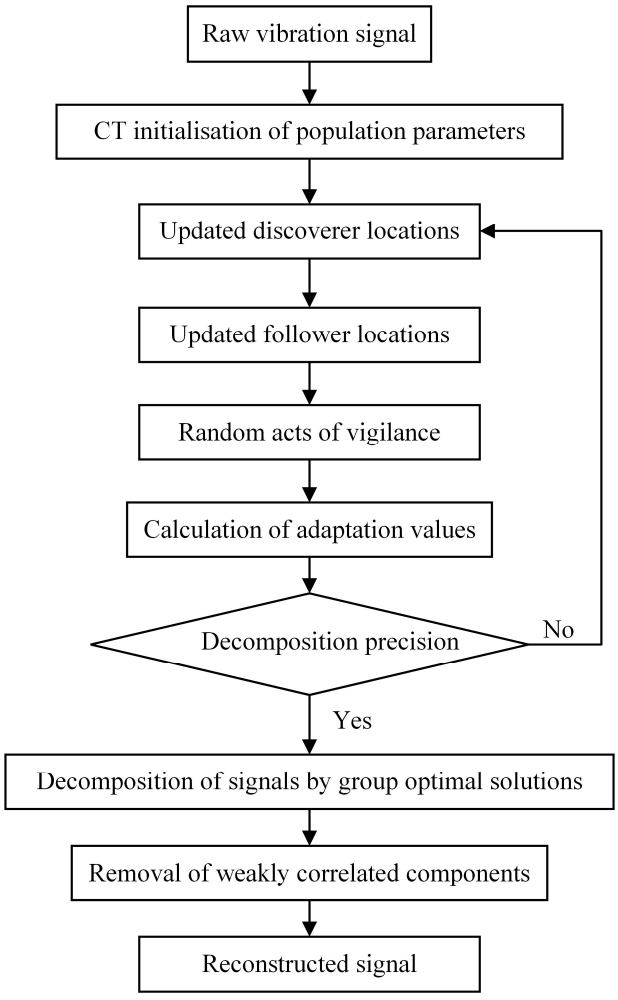


Figure 1. Flowchart of signal noise reduction based on AVMD algorithm.

3. An Enhanced Residual Neural Network Model Based on a Two-Dimensional Attention Mechanism (DAMResNet)

To improve the accuracy and generalization of the tool lifetime prediction model under variable working conditions, the one-dimensional residual block is enhanced using both channel and spatial attention mechanisms, forming a dual-dimensional attention mechanism. This section introduces the Dual Attention Mechanism Residual (DAMR) module, as shown in Figure 2.

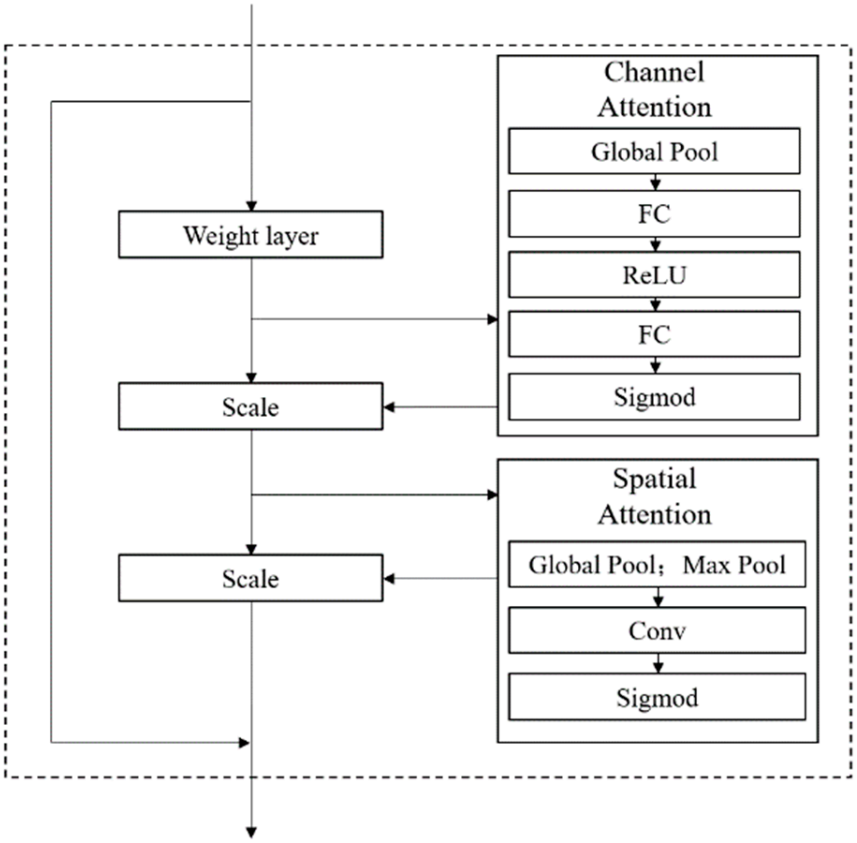


Figure 2. Structure diagram of Dual Attention Mechanism Residual module.

The DAMR structure consists of four components. The first component is the convolutional layer, where the input data undergo convolution, followed by batch normalization (BN) and ReLU activation. This produces a set of feature maps, with the height and width denoted by H and W, and the number of channels denoted by C.

The second component is the channel attention mechanism, which includes a global pooling layer, two fully connected layers, batch normalization (BN), ReLU activation, and a Sigmoid function. This mechanism learns to update the attention weights for each channel and generates a channel attention weight vector. The data are passed through the residual block to generate a set of feature maps, with height H, width W, and C channels. The module first compresses each channel of the input feature maps by performing global average pooling to compute the average across channels. This generates a vector of a length equal to the number of channels, representing the global average features for each channel, as shown in the following expression:

$$z_c = F_{sq}(u_c) = \frac{1}{H \cdot W} \sum_{i=1}^{H} \sum_{j=1}^{W} u_c(i,j) \tag{9}$$

where u_c denotes channel data in the feature map, z_c denotes a vector representing a channel after compression, and F_{sq} represents the compression operation.

Afterwards the compressed feature vectors are input to the fully connected layer to synthesize the channel feature values.

$$s_1 = F_{ex}(z, W) = \sigma(W_2\delta(W_1 z)) \tag{10}$$

where s_1 is the last obtained channel attention vector; W, W_1, and W_2 refer to the weight matrix; σ and δ are the Sigmod and Relu activation functions, respectively; and F_{ex} represents the vector operation performed.

The third part is the spatial attention mechanism structure, which contains two pooling layers, a convolutional layer, BN operation, and a Sigmoid function, and finally a weight matrix is obtained; the spatial attention mechanism structure is used to update the attention weights for the spatial location of the data to produce a weight matrix for spatial attention. The data will obtain a feature map containing multiple channels after passing through the residual block, the spatial attention mechanism will perform maximum pooling and average pooling operations on the values of all the channels for each feature point to obtain two pooling matrices, after that the two matrices will be spliced along the channel dimensions to obtain a feature map with different scaling information, and then a convolutional layer and a Sigmoid function will be used to control the weights to lie between 0 and 1.

$$s_2 = \sigma\left(\mathbf{W}\left[F_{avg}(u); F_{\max}(u)\right]\right) \tag{11}$$

where s_2 is the last obtained spatial attention vector, and F_{avg} and $F_{\max}$ represent the average pooling and maximum pooling operations.

The fourth part is the jump operation, the input data are weighted with the feature map obtained from the convolution part with the weight vector of the channel attention and the weight matrix of the spatial attention to obtain the training results of the three parts, the input data are summed up with the output of the training results, and finally the result of the summation is input into the activation function.

This section proposes a Dual Attention Mechanism Parallel Residual Network (DAM-ResNet), which combines an 18-layer residual network [22] with channel and spatial attention mechanisms to form an improved dual-dimensional attention model based on the characteristics of tool vibration signals. The model's input layer accepts a two-channel one-dimensional signal. The spatial attention mechanism is connected in series with the channel attention, and the residual structure incorporates the two-dimensional attention mechanism. This design ensures that input information is learned via convolution from the main channel while allowing cross-layer data propagation, preserving information integrity. The feature fusion layer, in conjunction with the fully connected layer, enables weighted learning of features from different sensors. The structure of DAMResNet is shown in Figure 3, and the layer architecture along with some parameter settings of the network are presented in Table 1.

Table 1. Algorithm parameter settings of DAMResNet.

Model Layer	Parameter	Output Dimension
Input	-	2048 × 1
Conv	64, 7	1024 × 64
Max Pooling	64, 3	1024 × 64
DAMR	64, 3	512 × 64
DAMR	64, 3	256 × 64

Table 1. *Cont.*

Model Layer	Parameter	Output Dimension
DAMR	128, 3	128 × 128
DAMR	128, 3	64 × 128
DAMR	256, 3	32 × 256
DAMR	256, 3	16 × 256
DAMR	512, 3	8 × 512
DAMR	512, 3	4 × 512
Average Pooling	-	512
Concatenate	-	1024
Dense	-	1

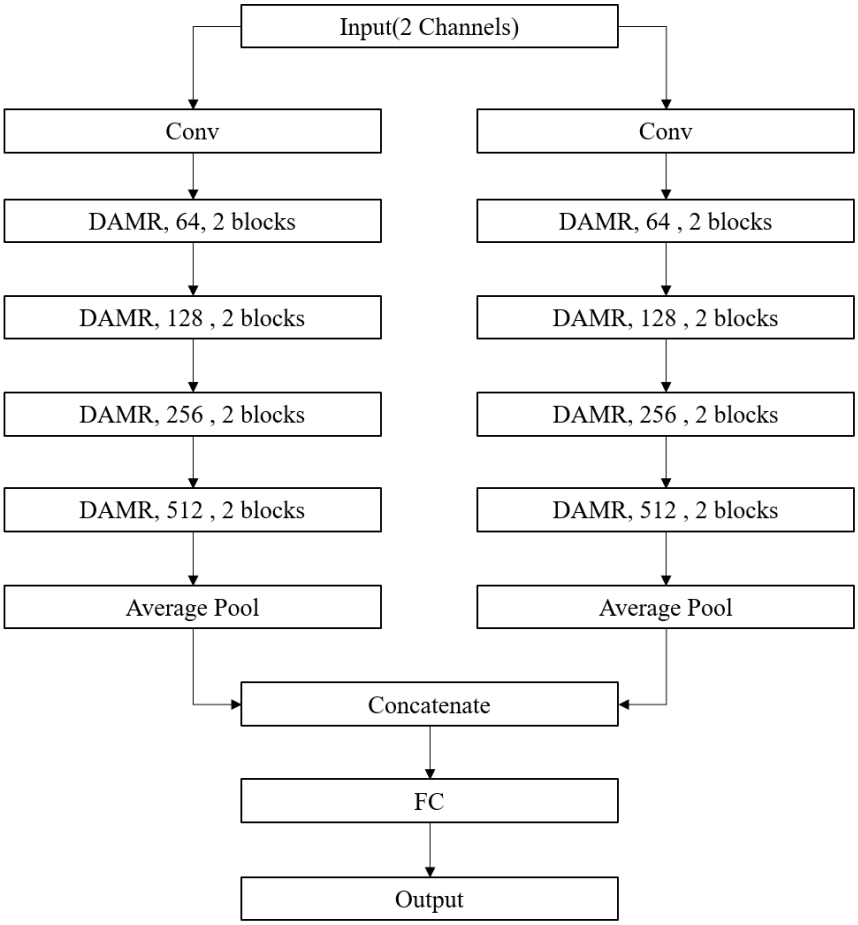


Figure 3. DAMResNet structure.

4. Tool Lifetime Prediction Method Based on AVMD-DAMResNet

This paper proposes a tool lifetime prediction method based on the AVMD-DAMResNet model, which integrates the AVMD noise reduction algorithm with DAMResNet. The method first performs noise reduction and signal reconstruction using the AVMD module, then feeds the reconstructed denoised signal into the DAMResNet model for feature learning, and finally predicts tool lifetime using a linear activation function. The framework of the AVMD-DAMResNet-based tool lifetime prediction method is shown in Figure 4.

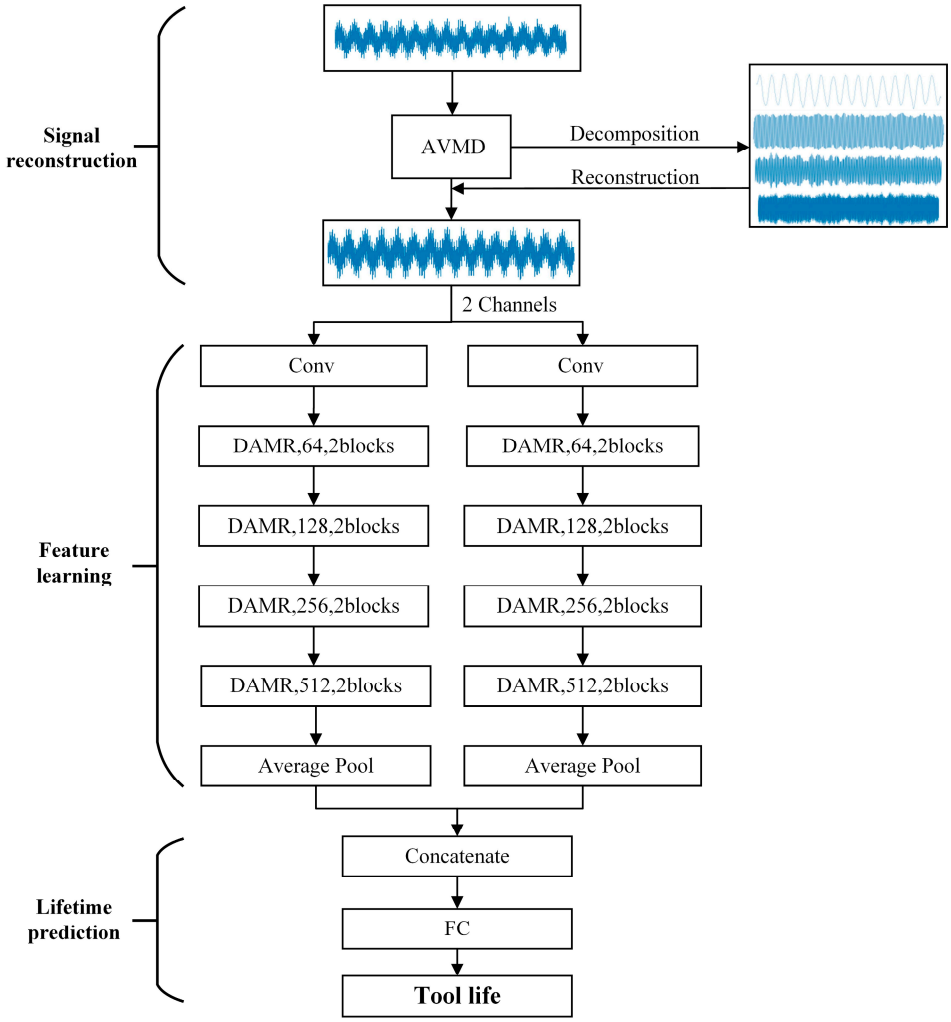


Figure 4. The proposed tool lifetime prediction method.

5. Experimental Validation

To verify the effectiveness of the AVMD-DAMResNet-based tool lifetime prediction method under noise interference and variable working conditions, this section conducts an ablation study on the adaptive noise reduction algorithm (AVMD) and the DAMResNet model, evaluating their performance in different machining environments. Section 5.1 uses simulated signals to mimic actual processing and evaluates the performance of the AVMD noise reduction algorithm under noise interference conditions. Section 5.2 builds a tool lifecycle signal acquisition platform based on the four-axis machining center TSIM-VMA8050V4 to conduct experiments under variable machining conditions, evaluating the prediction performance of the DAMResNet model. Section 5.3 assesses the tool lifetime prediction performance of the AVMD-DAMResNet method under variable working conditions.

5.1. Noise Interference Simulation Experiment

This experiment uses an analog signal as the original signal, with the frequencies of its components selected based on the data $f(t)$ obtained from the actual working condition signal acquisition experiment to validate the AVMD algorithm. The parameters of the analog signal $f(t)$ are as follows:

$$f(t) = 0.5\sin(30\pi t) + 0.5\sin(540\pi t) + 0.3\sin(750\pi t) + 0.5\sin(1500\pi t) \quad (12)$$

In this experiment, 1 dB, 5 dB, and 10 dB signal-to-noise ratio (SNR) noise is added to the analog signals, with a sampling time of 0.5 s and a sampling frequency of 6.4 kHz. The noisy signals are then processed for noise reduction using four methods: wavelet analysis, EMD, manually parameterized VMD, and the AVMD method. For wavelet analysis, Coif1 is selected as the wavelet basis function. In the manually parameterized VMD, the number of decomposition layers K is set to 5, and the penalty factor is set to 4800, calculated as 1.5 times the signal length. For signal reconstruction, components with a Spearman coefficient greater than 0.2 are selected after applying the four algorithms. The parameters of the AVMD algorithm are summarized in Table 2.

Table 2. Algorithm parameter settings.

Parameters	Value
Population size	40
Number of iterations	10
K	2–40
α	1000–20,000

Figure 5 illustrates the change in the algorithm's fitness during iteration. After repeated testing, the fitness value stabilizes at 5.865 after 10 iterations, which is considered the optimal result. Therefore, the number of decomposition layers K is set to 6, and the penalty factor is set to 5892.

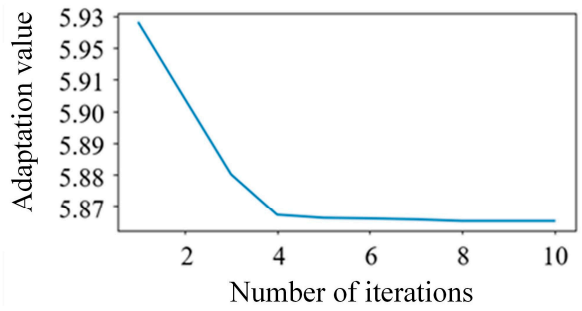


Figure 5. Fitness value iteration.

The signal-to-noise ratio (SNR) and Spearman's coefficient (ρ) are used as indicators to evaluate the performance of the reconstructed denoised signals. A higher SNR value indicates lower noise power and better noise reduction, while ρ represents the degree of correlation between the reconstructed signals and the original signals. Table 3 presents a comparison of the results from three algorithms: wavelet analysis, EMD, empirically parameterized VMD (EP-VMD), and the AVMD algorithm.

Table 3 shows that when the signal-to-noise ratio is 1 dB, the signal enhancement for wavelet analysis, EMD, empirically parameterized VMD, and AVMD are 3.642 dB, 4.271 dB, 3.999 dB, and 4.453 dB, respectively. The corresponding correlation coefficients are 0.7952, 0.8291, 0.8195, and 0.8369. Analyzing the data, it is evident that the noise reduction effect of AVMD outperforms the other three methods. Additionally, the higher correlation coefficient indicates that the AVMD algorithm exhibits superior noise reduction and anti-jamming capabilities under high-noise conditions, effectively preserving the original signal characteristics.

As an example, the analog signal $f(t)$ with an added signal-to-noise ratio (SNR) of 10 dB is used. Figure 6 presents the time–frequency diagram of the signal after decomposition using the proposed AVMD algorithm, while Table 3 lists the Spearman correlation coefficients of each component obtained from the decomposition.

Table 3. Decomposition effects of different methods.

Noise Added/dB	Decomposition Methods	SNR/dB	ρ
1 dB	Wavelet	4.632	0.7952
	EMD	5.271	0.8291
	EP-VMD	4.999	0.8195
	AVMD	5.453	0.8369
5 dB	Wavelet	7.032	0.8944
	EMD	7.711	0.9091
	EP-VMD	7.021	0.8732
	AVMD	8.077	0.9161
10 dB	Wavelet	10.174	0.9517
	EMD	11.267	0.9609
	EP-VMD	9.140	0.9357
	AVMD	12.149	0.9689

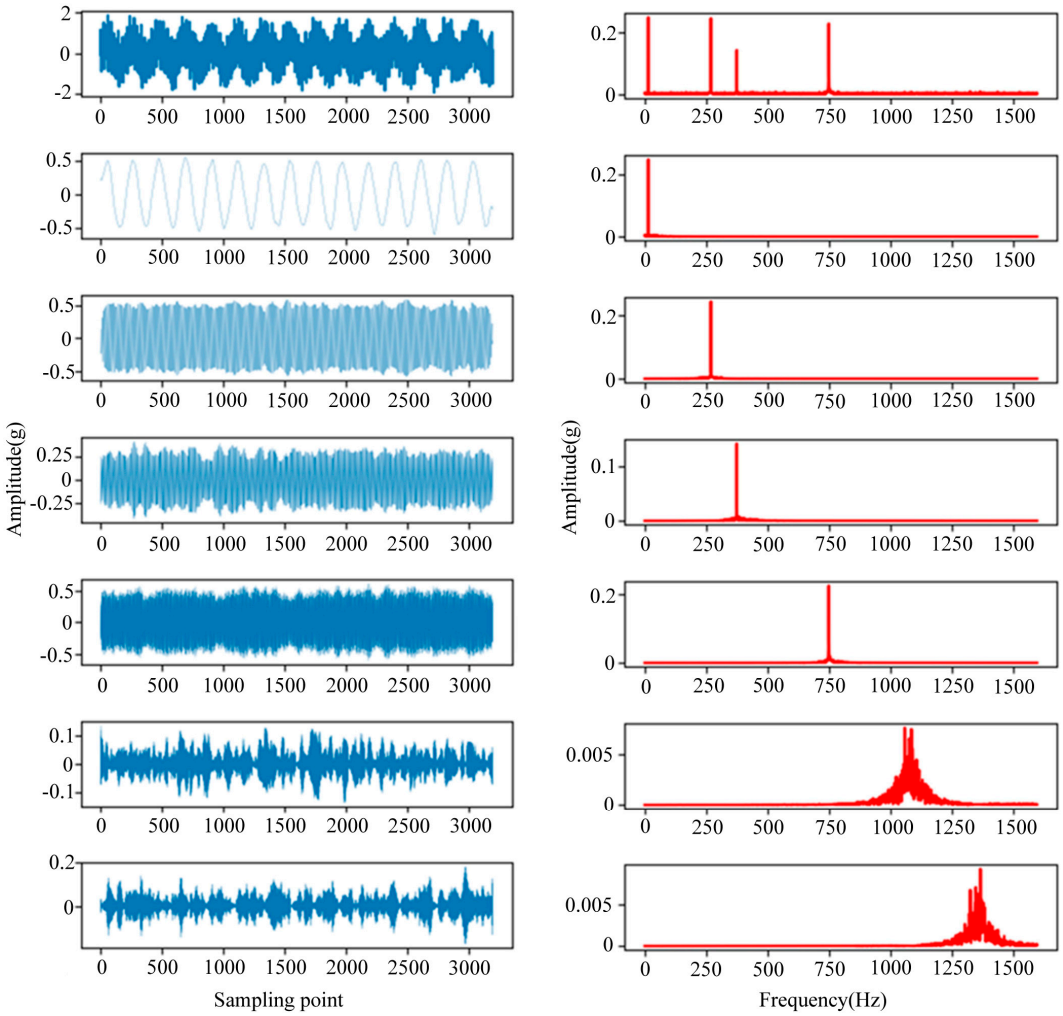


Figure 6. Time–frequency diagram of the simulated signal.

As shown in Figure 6, components 2 to 4 effectively separate the four main components of the analog signal after noise addition, while components 5 and 6 correspond to noise components with low amplitude. This demonstrates the strong noise reduction performance of the AVMD algorithm. From Table 4, the correlation coefficients of components 5 and 6 are both below 0.2, so these components are discarded in the reconstruction process. After reconstruction, the correlation coefficient of the denoised signal reaches 0.9689, indi-

cating that the reconstructed signal is highly consistent with the original vibration signal, demonstrating the high reliability of the AVMD algorithm.

Table 4. Correlation coefficient of each component.

Components	ρ
1	0.5177
2	0.5115
3	0.3155
4	0.5279
5	0.0948
6	0.1057

5.2. Tool Lifetime Prediction Experiment Under Variable Operating Conditions

The machining experiments under variable operating conditions were conducted on a four-axis machining center (TSIM-VMA8050V4). The tool selected was an M2Al high-speed steel end mill (four teeth, standard dimensions), and the milling mode was smooth milling). The workpiece material used was No. 45 steel, with a square shape and dimensions of 60 mm × 60 mm. Two spindle speeds, three feed speeds, and three sets of axial and radial depths of cut were selected for the milling experiments. The machining parameters are shown in Table 5. Figure 7 illustrates the experimental setup. The data acquisition system included an LC0108 piezoelectric acceleration sensor, an optical camera, and the NI-9230 and NI-cDAQ-9174 data acquisition cards. The two sensors, mounted at 90 degrees to the machine tool spindle using magnetic bases, had a sampling frequency of 6.4 kHz. The M2Al high-speed steel end mill was used for the planar milling of the workpiece, while the optical camera monitored the tool surface to measure tool wear and assess tool lifetime status.

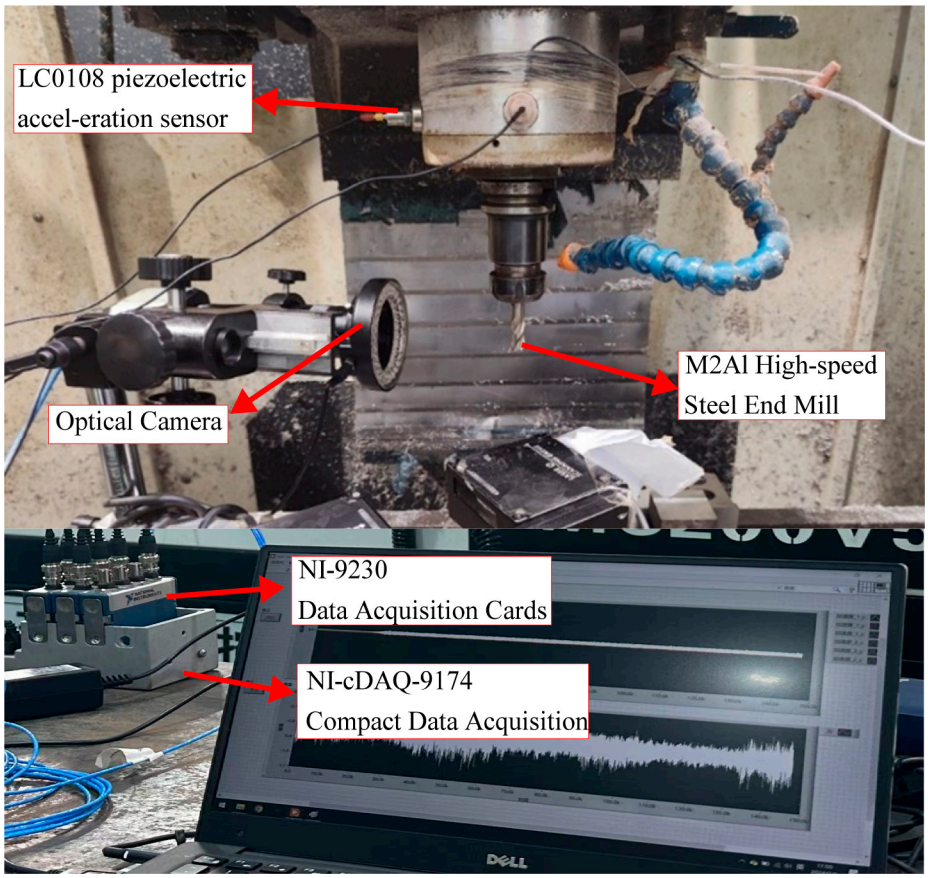


Figure 7. Experimental equipment.

Table 5. Experimental parameter settings.

Spindle Speed (r/min)	Feed Rate (mm/min)	Axial Depth of Cut (mm)	Radial Depth of Cut (mm)
1000	100	0.8	0.4
1000	110	1.0	0.5
1000	120	1.2	0.6
1100	100	1.0	0.6
1100	110	1.2	0.5
1100	120	0.8	0.4

The experimental data consisted of six milling cutters, with each cutter corresponding to one vibration signal dataset. Each dataset contained an average of 460 data samples, and each sample included two channels of vibration signals, each with 2048 sampling points. The wear value of the two edges of the cutter was used to determine whether the tool had reached its life standard. The tool lifetime for each sample was normalized to a range of 0 to 1, where 0 indicated no wear and 1 indicated that the wear criterion had been reached. This normalized value was used as the lifetime label. Table 6 provides a detailed description of the dataset samples.

Table 6. Dataset parameter.

Tool Number	Feed Rate (mm/min)	Spindle Speed (r/min)	Sample Size
C1	100	1000	504
C2	110	1000	432
C3	120	1000	392
C4	100	1100	480
C5	110	1100	456
C6	120	1100	456
Total			2720

5.2.1. Experiment on Tool Lifetime Prediction Based on DAMResNet Model Under Single Working Condition

In this subsection, the tool vibration signal dataset from Section 5.2 is used to train the DAMResNet model for tool lifetime prediction. Mean absolute percentage error (MAPE), root mean square error (RMSE), and the coefficient of determination (R^2) are selected to assess the model's performance. Root mean square error (RMSE) and the coefficient of determination (R^2) were chosen as the evaluation metrics for the lifetime prediction model.

$$\begin{aligned} MAPE &= \frac{1}{n}\sum_{i=1}^{n}\left(\left|y_p - y_r\right|/y_r\right) \\ RMSE &= \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(y_p - y_r\right)^2} \\ R^2 &= 1 - \frac{\sum_{i=1}^{n}\left(y_p - y_r\right)^2}{\sum_{i=1}^{n}\left(y_r - \overline{y_r}\right)^2} \end{aligned} \tag{13}$$

where n is the number of samples, y_p is the predicted value, and y_r is the actual value.

Figure 8 shows the test results of the DAMResNet model on the C1, C3, and C5 datasets. The black curve represents the tool's lifetime after sample normalization, the blue curve shows the DAMResNet model predictions, and the blue bar indicates the absolute

value of the lifetime prediction error. It can be concluded that the DAMResNet model accurately predicts the tool's lifetime across different datasets, with higher accuracy during the normal wear stage.

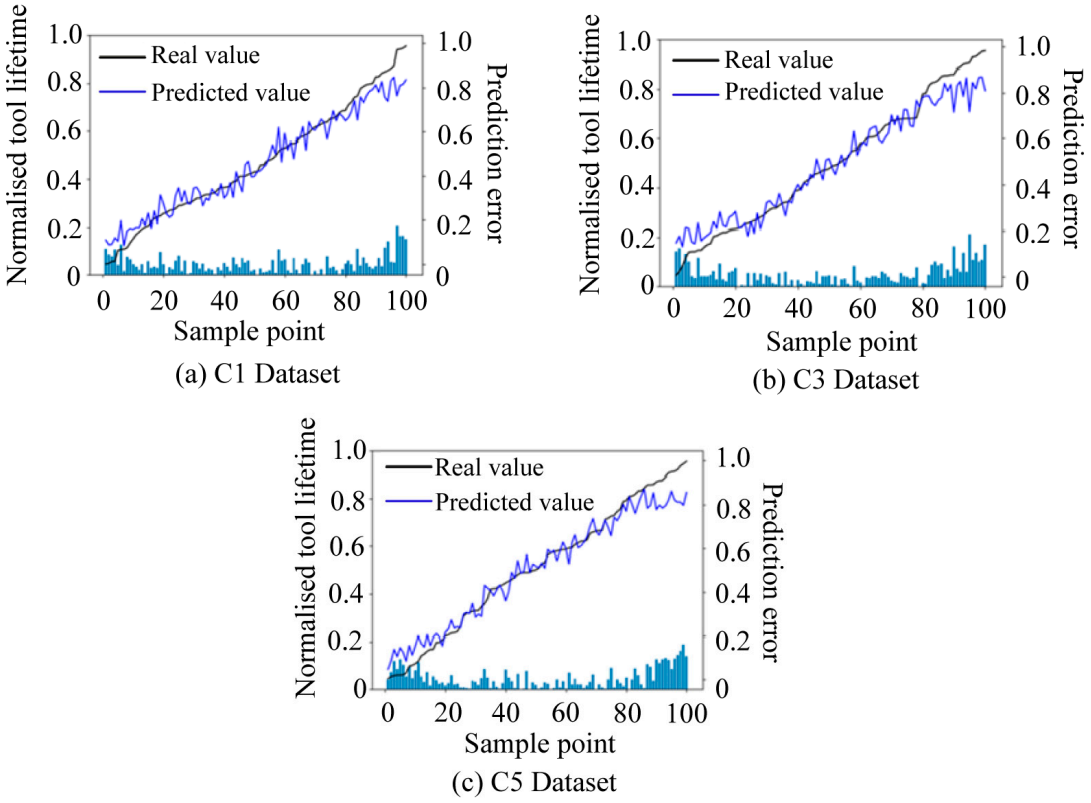


Figure 8. Prediction results of DAMResNet.

To verify the effectiveness of the DAMResNet model, comparison experiments were conducted using the DAMResNet model and other algorithms: a one-dimensional convolutional network (1D-CNN) [23], a bi-directional long short-term memory network (BiLSTM) [24], a primitive residual network (ResNet) [25], and the residual network optimized with a channel attention mechanism (SE-ResNet) [26]. Table 7 presents the prediction performance of the different models.

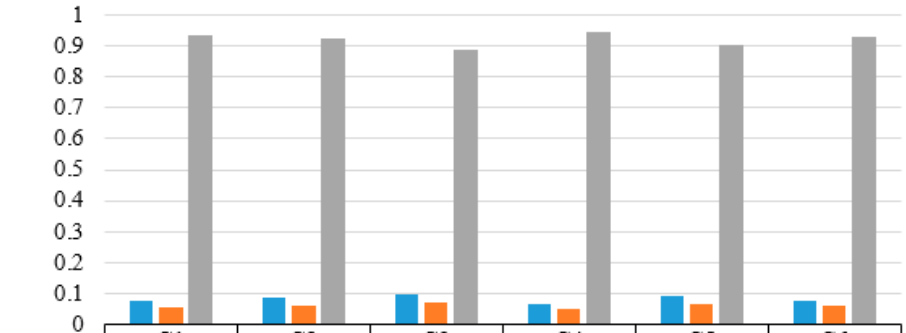
Table 7. Comparison of different models under single working condition.

Models	RMSE	MAPE	R^2
DAM-PResNet	0.061	0.047	0.944
1D-CNN	0.087	0.065	0.899
BiLSTM	0.098	0.072	0.873
ResNet	0.088	0.064	0.908
SE-ResNet	0.076	0.055	0.928

As shown in Table 7, the R^2 value of the DAMResNet model exceeds 0.94, demonstrating the best prediction performance among all models. Compared to the unimproved ResNet model, the DAMResNet model achieved a 29.3% reduction in RMSE, a 26.6% reduction in MAPE, and a 3.9% increase in R^2. These improvements demonstrate that the DAMResNet model reduces prediction errors and enhances accuracy by integrating channel and spatial feature learning.

5.2.2. Experiments on Tool Lifetime Prediction Based on DAMResNet Model Under Variable Working Conditions

To evaluate the generalization ability of the DAMResNet model, the trained model, initially developed on a single working condition dataset, was tested on datasets from other conditions, as shown in Figure 9. The C1 to C6 datasets served as the training set, while the remaining datasets were used for testing.

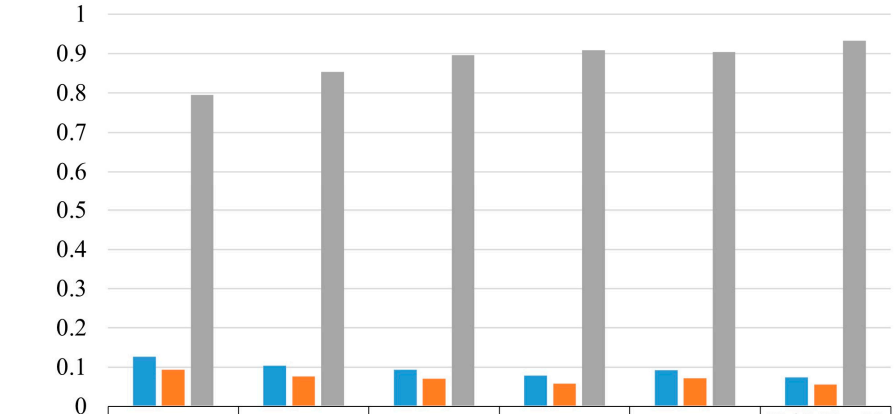


	C1	C2	C3	C4	C5	C6
RMSE	0.074	0.084	0.095	0.067	0.089	0.076
MAPE	0.055	0.061	0.068	0.051	0.065	0.058
R2	0.933	0.923	0.886	0.944	0.902	0.931

Figure 9. Prediction results of the DAMResNet model with different datasets.

As shown in Figure 9, the DAMResNet model, trained on a single dataset, maintains good prediction accuracy for tool lifetime under different working conditions. Although performance decreases compared to the single working condition, the R^2 remains above 0.92 on average across all datasets. These results demonstrate that the DAMResNet model has strong generalization ability and is valuable for real-world machining applications.

To evaluate the performance improvement of the DAMResNet model compared to other tool lifetime prediction algorithms, comparison experiments were conducted using a CNN [27], BiLSTM, ResNet, SE-ResNet, and ConvLSTM [28] as baseline models. The C1 dataset was used for training, and the remaining datasets were used for testing. Figure 10 shows the prediction results of the different models.



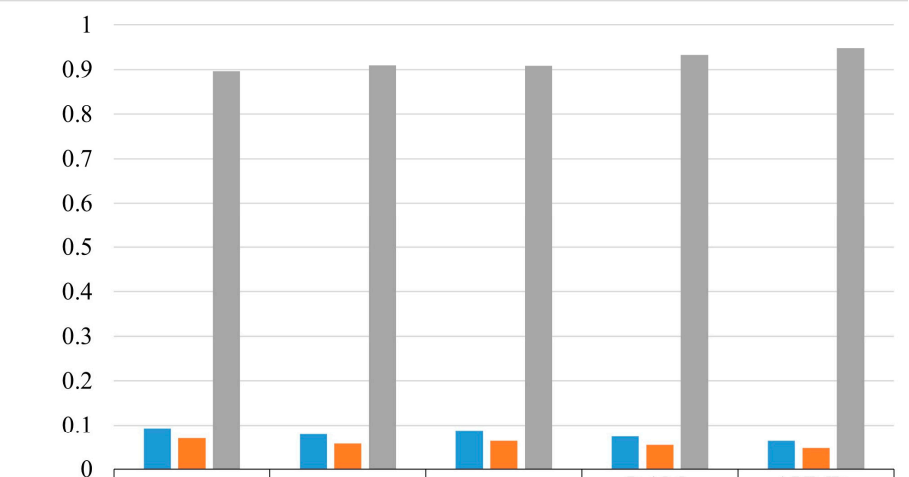
	CNN	BiLSTM	ResNet	SE-Resnet	ConvLSTM	DAMResNet
RMSE	0.127	0.104	0.093	0.079	0.092	0.074
MAPE	0.093	0.076	0.07	0.058	0.072	0.055
R2	0.795	0.853	0.896	0.909	0.904	0.933

Figure 10. Prediction results of the DAMResNet model compared to other tool lifetime prediction models.

As shown in Figure 10, the DAMResNet model outperforms several other models in terms of prediction accuracy and error reduction. Compared to the original ResNet model, the DAMResNet model shows a 4.1% improvement in R^2, a 22.9% reduction in RMSE, and a 23.7% reduction in MAPE. These improvements demonstrate that the incorporation of the two-dimensional attention mechanism and parallel structure strengthens the model's ability to extract tool lifetime features and enhances its generalization ability.

5.3. Experiments on AVMD-DAMResNet-Based Tool Lifetime Prediction Under Variable Working Conditions

The C1 dataset was used for training, and the C2–C5 datasets were used for testing. The Channel Attention Mechanism Residual Model (SE-ResNet), Single-Channel Dual-Dimension Attention Mechanism Residual Model (DAM-ResNet), and DAM-PResNet were compared with the AVMD-DAMResNet model proposed in this paper. The experimental results are shown in Figure 11.



	ResNet	SE-ResNet	DAM-ResNet	DAM-PResNet	AVMD-DAMResNet
RMSE	0.093	0.079	0.088	0.074	0.064
MAPE	0.07	0.058	0.064	0.055	0.048
R2	0.896	0.909	0.908	0.933	0.948

Figure 11. Evaluation results of different models under variable operating conditions.

A comparative analysis of Figures 10 and 11 shows that the predictive performance of the DAMResNet model is further enhanced by incorporating the AVMD noise reduction module. Compared to the original ResNet, the AVMD-DAMResNet model improves R^2 by 5.8%, reduces RMSE by 31.2%, and lowers MAPE by 31.4%. Compared to the DAMResNet model without the noise reduction module, the AVMD-DAMResNet model improves R^2 by 1.5%, reduces RMSE by 13.5%, and decreases MAPE by 12.7%. The experimental results confirm that the AVMD-DAMResNet model is well-suited for accurate tool lifetime prediction under complex machining conditions.

6. Conclusions

This paper addresses the limitations of tool lifetime prediction methods under noise interference and variable operating conditions by focusing on two key aspects: effective noise reduction and improved generalization of lifetime prediction models. A tool lifetime prediction method based on AVMD-DAMResNet is proposed. The main contributions are summarized as follows:

(1) This method proposes an adaptive variational modal decomposition (AVMD)-based algorithm to address the issue of reduced model prediction accuracy caused by vibration signal noise. The adaptive selection of parameters for the variational modal noise reduction algorithm is achieved using the improved sparrow search algorithm, enhancing the model's noise reduction performance.

(2) This method introduces a parallel residual network tool prediction model (DAMResNet) based on an improved two-dimensional attention mechanism to address the performance degradation of tool lifetime prediction models under variable operating conditions. The residual module in the deep residual network is enhanced with channel and spatial attention mechanisms. The model's feature weighting is optimized, and a parallel structure is used to separately learn two groups of signal features, fuse multi-channel features, and accurately predict tool lifetime.

(3) By combining the AVMD noise reduction algorithm with the DAMResNet model, the AVMD-DAMResNet tool lifetime prediction method is developed, which predicts tool lifetime under noise interference and variable operating conditions. The experimental results demonstrate that the proposed method improves R^2 by 5.8%, reduces RMSE by 31.2%, and lowers MAPE by 31.4% compared to the original ResNet model, achieving the best prediction performance among various improved ResNet models.

The tool lifetime prediction model proposed in this paper requires a sufficient number of labeled samples during training. However, in practical engineering applications, obtaining diverse tool data and labeling them is both costly and time-consuming. Additionally, the model demands more training samples under variable working conditions than under constant conditions. Therefore, future work will focus on enhancing the model's generalization performance while improving its prediction accuracy under small sample conditions.

Author Contributions: J.K.: writing—review and editing, writing—original draft, visualization, validation, software, resources, and methodology; T.W.: supervision, project administration, and funding acquisition; Y.L.: investigation, data curation, and conceptualization; Y.W.: software and writing—review and editing; Y.Z.: software and validation; Y.T.: project administration and funding acquisition. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported by the National Key Research and Development Program (grant number 2022YFB3303601).

Data Availability Statement: The data will be made available by the authors on request.

Conflicts of Interest: The authors have no competing interests to declare that are relevant to the content of this article.

References

1. Wu, J.; Wang, J.; Chen, H. A Method for Predicting Tool Remaining Useful Life: Utilizing BiLSTM Optimized by an Enhanced NGO Algorithm. *Mathematics* **2024**, *12*, 2404. [CrossRef]
2. Jemielniak, K.; Urbański, T.; Kossakowska, J.; Bombiński, S. Tool condition monitoring based on numerous signal features. *Int. J. Adv. Manuf. Technol.* **2012**, *59*, 73–81. [CrossRef]
3. Wang, M.; Wang, J. CHMM for tool condition monitoring and remaining useful life prediction. *Int. J. Adv. Manuf. Technol.* **2012**, *59*, 463–471. [CrossRef]
4. Guo, R.; Wang, Y.; Zhang, H.; Zhang, G. Remaining useful life prediction for rolling bearings using EMD-RISI-LSTM. *IEEE Trans. Instrum. Meas.* **2021**, *70*, 3509812. [CrossRef]
5. Sayyad, S.; Kumar, S.; Bongale, A.; Kotecha, K.; Abraham, A. Remaining Useful-Life Prediction of the Milling Cutting Tool Using Time–Frequency-Based Features and Deep Learning Models. *Sensors* **2023**, *23*, 5659. [CrossRef]
6. Benkedjouh, T.; Zerhouni, N.; Rechak, S. Tool wear condition monitoring based on continuous wavelet transform and blind source separation. *Int. J. Adv. Manuf. Technol.* **2018**, *97*, 3311–3323. [CrossRef]
7. Sun, L.; Zhao, C.; Huang, X.; Ding, P.; Li, Y. Cutting tool remaining useful life prediction based on robust empirical mode decomposition and Capsule-BiLSTM network. *Proc. Inst. Mech. Eng. Part C J. Mech. Eng. Sci.* **2023**, *237*, 3308–3323. [CrossRef]

8. Liu, R.; Meng, G.; Yang, B.; Sun, C.; Chen, X. Dislocated time series convolutional neural architecture: An intelligent fault diagnosis approach for electric machine. *IEEE Trans. Ind. Inform.* **2016**, *13*, 1310–1320. [CrossRef]
9. Benkedjouh, T.; Medjaher, K.; Zerhouni, N.; Rechak, S. Health assessment and life prediction of cutting tools based on support vector regression. *J. Intell. Manuf.* **2015**, *26*, 213–223. [CrossRef]
10. Yuan, J.; Liu, L.; Yang, Z.; Zhang, Y. Tool wear condition monitoring by combining variational mode decomposition and ensemble learning. *Sensors* **2020**, *20*, 6113. [CrossRef]
11. Yuan, J.; Liu, L.; Yang, Z.; Bo, J.; Zhang, Y. Tool wear condition monitoring by combining spindle motor current signal analysis and machined surface image processing. *Int. J. Adv. Manuf. Technol.* **2021**, *116*, 2697–2709. [CrossRef]
12. Liu, C.; Zhang, L.; Niu, J.; Yao, R.; Wu, C. Intelligent prognostics of machining tools based on adaptive variational mode decomposition and deep learning method with attention mechanism. *Neurocomputing* **2020**, *417*, 239–254. [CrossRef]
13. Sun, M.; Han, Y.; Guo, K.; Sivalingam, V.; Huang, X.; Sun, J. A milling tool wear predicting method with processing generalization capability. *J. Manuf. Process.* **2024**, *120*, 975–1001. [CrossRef]
14. Zhao, R.; Wang, D.; Yan, R.; Mao, K.; Shen, F.; Wang, J. Machine health monitoring using local feature-based gated recurrent unit networks. *IEEE Trans. Ind. Electron.* **2017**, *65*, 1539–1548. [CrossRef]
15. Zhao, R.; Yan, R.; Chen, Z.; Mao, K.; Wang, P.; Gao, R.X. Deep learning and its applications to machine health monitoring. *Mech. Syst. Signal Process.* **2019**, *115*, 213–237. [CrossRef]
16. Qiao, H.; Wang, T.; Wang, P. A tool wear monitoring and prediction system based on multiscale deep learning models and fog computing. *Int. J. Adv. Manuf. Technol.* **2020**, *108*, 2367–2384. [CrossRef]
17. Zhou, Y.; Zhi, G.; Chen, W.; Qian, Q.; He, D.; Sun, B.; Sun, W. A new tool wear condition monitoring method based on deep learning under small samples. *Measurement* **2022**, *189*, 110622. [CrossRef]
18. Kumar, M.P.; Dutta, S.; Murmu, N. Tool wear classification based on machined surface images using convolution neural networks. *Sādhanā* **2021**, *46*, 130. [CrossRef]
19. Duan, J.; Shi, T.; Zhou, H.; Xuan, J.; Wang, S. A novel ResNet-based model structure and its applications in machine health monitoring. *J. Vib. Control* **2021**, *27*, 1036–1050. [CrossRef]
20. Liu, Q.; Zhang, Y.; Li, M.; Zhang, Z.; Cao, N.; Shang, J. Multi-UAV path planning based on fusion of sparrow search algorithm and improved bioinspired neural network. *IEEE Access* **2021**, *9*, 124670–124681. [CrossRef]
21. He, Y.; Wang, M. Dynamic step opposition-based learning sparrow search algorithm for UAV path planning. *Clust. Comput.* **2025**, *28*, 44. [CrossRef]
22. Ma, X.; Wang, Y.; Qin, J.; Wang, Z.; Liu, Z. A bearing fault diagnosis model with convolutional cross transformer and ResNet18. *Meas. Sci. Technol.* **2024**, *36*, 016132. [CrossRef]
23. Yin, Y.; Wang, S.; Zhou, J. Multisensor-based tool wear diagnosis using 1D-CNN and DGCCA. *Appl. Intell.* **2023**, *53*, 4448–4461. [CrossRef]
24. Wu, X.; Li, J.; Jin, Y.; Zheng, S. Modeling and analysis of tool wear prediction based on SVD and BiLSTM. *Int. J. Adv. Manuf. Technol.* **2020**, *106*, 4391–4399. [CrossRef]
25. Guan, R.; Cheng, Y.; Jin, Y.; Zhou, S.; Gai, X.; Lu, M. Research on tool wear prediction for milling high strength steel based on DenseNet-ResNet-GRU. *J. Mech. Sci. Technol.* **2024**, *38*, 3585–3596. [CrossRef]
26. Wu, Z.; He, L.; Wang, W.; Ju, Y.; Guo, Q. A Fault Prediction Method for CNC Machine Tools Based on SE-ResNet-Transformer. *Machines* **2024**, *12*, 418. [CrossRef]
27. Ambadekar, P.; Choudhari, C. CNN based tool monitoring system to predict life of cutting tool. *SN Applied Sciences* **2020**, *2*, 860. [CrossRef]
28. Deng, F.; Chen, Z.; Liu, Y.; Yang, S.; Hao, R.; Lyu, L. A novel combination neural network based on convlstm-transformer for bearing remaining useful life prediction. *Machines* **2022**, *10*, 1226. [CrossRef]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.