sensors

MDPI

Article

Tool Wear Monitoring in Milling Based on Fine-Grained Image Classification of Machined Surface Images

Jing Yang, Jian Duan, Tianxiang Li, Cheng Hu, Jianqiang Liang and Tielin Shi *

School of Mechanical Science and Engineering, Huazhong University of Science and Technology, 1037 Luoyu Road, Wuhan 430074, China
* Correspondence: tlshi@hust.edu.cn

Abstract: Cutting tool wear state assessment during the manufacturing process is extremely significant. The primary purpose of this study is to monitor tool wear to ensure timely tool change and avoid excessive tool wear or sudden tool breakage, which causes workpiece waste and could even damage the machine. Therefore, an intelligent system, that is efficient and precise, needs to be designed for addressing these problems. In our study, an end-to-end improved fine-grained image classification method is employed for workpiece surface-based tool wear monitoring, which is named efficient channel attention destruction and construction learning (ECADCL). The proposed method uses a feature extraction module to extract features from the input image and its corrupted images, and adversarial learning is used to avoid learning noise from corrupted images while extracting semantic features by reconstructing the corrupted images. Finally, a decision module predicts the label based on the learned features. Moreover, the feature extraction module combines a local cross-channel interaction attention mechanism without dimensionality reduction to characterize representative information. A milling dataset is conducted based on the machined surface images for monitoring tool wear conditions. The experimental results indicated that the proposed system can effectively assess the wear state of the tool.

Keywords: tool wear; tool condition monitoring; machined surface images; fine-grained image classification; channel attention; convolutional neural network (CNN)

check for updates

Citation: Yang, J.; Duan, J.; Li, T.; Hu, C.; Liang, J.; Shi, T. Tool Wear Monitoring in Milling Based on Fine-Grained Image Classification of Machined Surface Images. *Sensors* **2022**, *22*, 8416. https://doi.org/10.3390/s22218416

Academic Editor: Roberto Teti

Received: 9 October 2022
Accepted: 31 October 2022
Published: 2 November 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Copyright: © 2022 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

The rapid development of advanced manufacturing is closely related to a country's economy [1]. Milling is an indispensable part of modern manufacturing industries and optimizing its machining process is crucial to achieving substantial economic benefits and improving product quality. Manufacturing cost is mainly affected by the cutting tool, power, machining efficiency, and machined surface quality [2]. Cutting tool wear condition is an important factor during the manufacturing process, which directly affects the quality of products and the operation of equipment. Moreover, monitoring tool wear condition is necessary to ensure timely tool change; an early tool replacement increases cost, while a tool replaced too late causes reduced workpiece quality and even damages the machine tool. In the machining process, the total processing cost of cutting tools and replacing them ranges from 3% to 12% [3]; at the same time, the downtime caused by cutting tools can represent up to 20% of total machining downtime [4]. Tool wear is a normal phenomenon because the tool is in contact with the chip and workpieces. Moreover, the tools undergoes various failures mechanisms, such as adhesion, abrasion, chipping, diffusion, and plastic deformation. Due to the highly complex physics behind tool wear, it is challenging to avoid tool wear. Therefore, tool wear monitoring is essential in advanced manufacturing to optimize the machining process and maintain the quality of the manufactured product. For these purposes, it is crucial to develop an automated and accurate tool wear monitoring system to generate warnings for tool wear, which has been the focus of various research studies in tool wear monitoring.

Recently, cutting tool condition monitoring has focused chiefly on extracting degradation information by various physical properties, and the research methodologies have been divided into the following categories [5]: direct approaches, in which a machine vision system as a common method [6,7] is used to directly measure the areas of tool wear for evaluating tool conditions; and indirect approaches, in which the wear state is evaluated by analyzing signals from various sensors, such as machining force [8], vibration [9], temperature [10], acoustic emissions (AE) [11], and machined surface roughness [12–17]. Surface texture images, as relatively easily obtained monitoring data, can be extracted by cost-effective data acquisition devices to provide rich information for diagnosis and monitoring applications. The machined surfaces' texture not only directly realizes the evaluation of machined surface quality but is also indirectly utilized to reflect various types of tool wear, such as flank wear, crater wear, nose wear, fracture, and breakage. Compared with the abovementioned methods, the most notable advantages of analyzing machined surface images for monitoring tool wear states are non-invasive, low-cost, and flexible technique.

The machined surface images could provide rich geometrical characteristics for diagnosing the soundness of the state of cutting tools [18]. The massive and unstructured raw image data brings new opportunities and challenges to vison-based tool condition monitoring. Because machine vision-based tool condition monitor methods can learn the texture characteristics of an image from massive image data and automatically build the corresponding monitoring models, these have attracted increasing attention in recent vision-based tool condition monitoring studies. The machine vision approach relies on handcrafted feature design to extract sensitive degradation features using prior knowledge and expertise from the acquired monitoring data. Then, the extracted features are sent into machine learning models to assess the target values. At present, there is a rich research literature on the topic of machine learning surface texture inspection for tool wear assessment. Bhat et al. [12] presented a support vector machine (SVM) model to predict the state of tools using features extracted from the grey level co-occurrence matrix (GLCM) of machined surface images. Dutta et al. [13,14] proposed a support vector machine-based regression model to assess tool flank wear using extracted features from turned surface images. Kassim et al. [15] developed a run-length statistical method to monitor the tool condition based on the machined surface images by using a machined vision technique. Riego et al. [16] designed an extremely randomized tree algorithm to monitor the wear state by means of supervised classification. Li et al. [17] studied a micro-vision system to monitor the wear of a cutting tool, which combines an insert image with workpiece texture. These methods aim to develop an intelligent method to monitor the level of tool wear through the machined surface. Although these automated monitoring systems employed handcrafted feature extraction by machine vision to achieve decent results for monitoring tool wear, which requires significant computational effort and related knowledge, they are less intelligent.

To help enhance the intelligence capability of tool monitoring systems, a new branch of artificial intelligence called deep learning [19] has widely emerged in tool condition detection fields and has provided excellent results. Deep learning is a particular kind of machine learning structure with more powerful feature learning abilities, including deep belief networks (DBNs), recurrent neural networks (RNNs), and convolutional neural network (CNNs). It achieves advanced accuracy in various tasks, such as computer vision [20], speech recognition [21], and natural language processing [22]. Compared with traditional handcrafted feature extraction, deep learning-based tool wear monitoring studies [23–28] realized the complex correlation between the input (signal data) and the target output values (tool wear). These papers observed that the latent features are automatically learned from raw data for deep learning technology, which is advantageous over manually designed features. Although deep learning-based tool condition monitoring has been developed in recent years, very few studies have focused on deep learning-based machined surface monitoring methods. In 2021, Kumar et al. [29] proposed a deep CNN architecture

for intelligent wear monitoring of a cutting tool using machined surface images in turning. The model applied preprocessing images as an input to overcome inhomogeneous equalization. Although the model has achieved promising monitoring results, the preprocessing operation is not intelligent enough. In this paper, an end-to-end intelligence method should be proposed for tool condition monitoring, which creates direct mapping from raw data (machined surface images) to the needed results (the level of cutting wear). The workpiece textures with different tool wear states are characterized by visually shared global texture structure. The wear category is to be recognized by local regions with smaller visual differences. Fine-grained image classification, as a challenging task in computer vision, can distinguish different local details under the same global structure. The discriminative local regions play a critical factor in fine-grained classification and are identified to classify the target categories. Therefore, fine-grained classification has been introduced to recognize similar workpiece textures to identify the tool wear state.

The fine-grained image classification method has been widely applied in various classification tasks with the same global structure (such as animals, product brands, and vehicles) [30–35]. Deconstruction and construction learning (DCL) [36] is a newly proposed fine-grained classification model and differs from traditional classification methods with significant differences between categories, which aim to extract discriminative features from similar global structures. The architecture for DCL includes both destruction and construction streams, where the former is used to enhance recognition robustness while the latter is used to simulate the semantic correlations between image regions. As a result, DCL, as an end-to-end approach, may be readily learned without manual intervention. It, in particular, is light in weight, fast in reasoning, and practical. In this study, an intelligent method was proposed focusing on cutting tool wear monitoring from machined surface images based on the architecture of a CNN with fine-grained classification.

An attention mechanism is used to ensure that most salient information is noticed, which has been proven effective in many previous studies [37–39]. Hu et al. [40] developed the squeeze-and-excitation (SE) module to learn channel features by using interchannel relationships. However, describing the learned features by dimensionality reduction is suboptimal. Wang et al. [41] presented an efficient channel attention (ECA) module as an improvement of the squeeze-and-excitation module.

In this article, efficient channel attention destruction-construction learning (ECADCL) is proposed to monitor the tool wear state (sharp, normal, or dull) from machined surface images. The proposed ECADCL, including the feature extraction module, destruction-construction module, and decision module, combines the conventional CNN and the idea of DCL for identifying the local detail from a similar global structure. Considering the ability to learn feature representation, the ECA module is introduced into the feature extraction module and named ECACNN. Ultimately, the experimental results confirm that ECADCL offers better tool wear monitoring performance than the existing methods.

The main contents of this study are organized as follows.

(1) The architecture of ECACNN is structured by combining a typical CNN and attention mechanisms. This architecture can enhance the channel relationships of features and extract efficient information from texture images.
(2) Based on fine-grained image recognition, ECADCL is designed to enhance the representation of local details.
(3) An automatic monitoring system, which is a combination of the designed algorithms and the image acquisition system, is designed to detect the tool conditions based on machined surface images.
(4) The experiments verify that the proposed method is accurate and effective. Compared with handcrafted feature extraction methods and conventional CNN methods, we show that the proposed ECADCL can obtain competitive performance with handcrafted feature extraction and solve local feature recognition problems compared to conventional CNNs.

The rest of this article is organized as follows. In Section 2, the experimental system is described. In Section 3, the proposed ECADCL technique is elaborated. Section 4 presents the experimental results. Section 5 concludes the article.

2. System Setup

The experimental data used in this article are from a TC500 high-speed CNC milling machine, which uses uncoated carbide end mill inserts (ZMCC PML-4E-D6) with a four-insert-type end mill cutter 6 mm in diameter. The work material used when collecting the data was SUS 316 steel. The parameter combinations for the conducted experiments are listed in Table 1.

Table 1. Machining parameters used in the experiments.

Equipment Model	TC500
Spindle speed	6000 rpm
Cutting depth	0.5 mm
Cutting width	2.0 mm
Feed rate	1500 mm/min
Workpiece material	SUS 316 steel
Tool type	ZMCC PML-4E-D6

In the experiments, with the goal of minimizing the complexity of designing computer vision algorithms, using a camera precisely acquires a high-quality image of the workpiece surface. We created an image acquisition system that included a camera and an illuminator to capture images of all machined surfaces. Figure 1 illustrates that a complementary metal-oxide-semiconductor (CMOS) industrial camera (DMK 33GX183), a coaxial light source, and a workpiece are the key components of the image acquisition system. The CMOS camera and coaxial light source were mounted in a vertical manner, and the relative distance between the workpiece and camera was captured at certain predefined positions. The surfaces of the resulting machined workpiece were collected as 5472 × 3648-pixel red-green-blue (RGB) digital images by the camera, which was connected to a computer equipped with image acquisition capabilities, at five different positions relative to the machined surface.

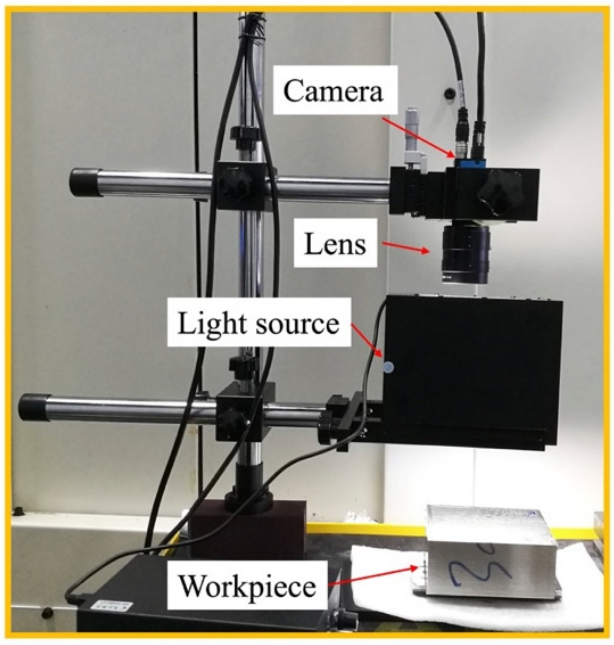


Figure 1. Prototype of the machined surface image acquisition system.

In our study, we used the average tool flank wear ($VB_{average}$) as the measurement standard, while the value of $VB_{average}$ was measured using a Keyence VHX-700F microscope. Flank wear appears to be caused by the wear of the cutter edge, which will affect the surface texture wear value. Thus, according to the tool wear value, the evaluation criteria of tool condition can be divided into three categories: sharp ($VB_{average}$ < 50 μm), regular (50 μm < $VB_{average}$ < 80 μm), and dull ($VB_{average}$ > 80 μm), which are shown in Figure 2. The surface image textures are divided into fine-grained curve texture and coarse-grained linear texture.

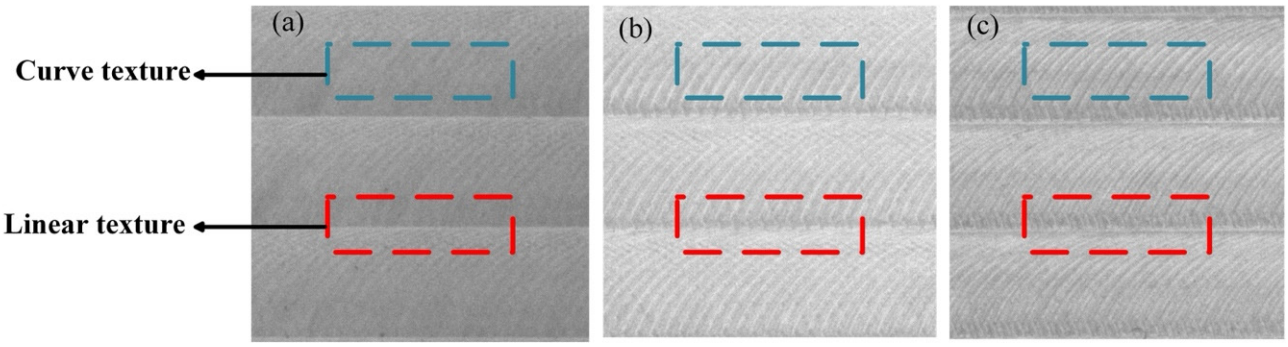


Figure 2. Image samples corresponding to different tool wear states: (**a**) sharp; (**b**) normal; and (**c**) dull.

3. Methodology

In this work, cutting wear monitoring based on the machined surface is regarded as an image classification problem. We aim to design an end-to-end intelligent model for tool wear monitoring that directly maps the input data to the target output. The model is represented as $C(\mathcal{X}, \mathcal{Y} \mid \theta)$, where $\mathcal{X}$ represents a set of input images, $\mathcal{Y}$ denotes the label of $\mathcal{X}$, and θ denotes all the learnable parameters. It should be noted that the different levels of wear on machined surface images have the same global structure and differ only in specific local details. In order to find the discriminating area among different tool wear states, the fined-grained method is introduced for realizing the monitoring tool condition. The procedure of this method is shown in Figure 3. Basically, it includes three stages: acquisition, training, and inference. The first part consists of data acquisition and data processing. In the second part, the proposed method applied the channel attention mechanism and the destruction-construction structure, and the training process mainly consists of feature extraction with attention mechanism, destruction learning (DL), and construction learning (CL). In the destruction process, the region confusion mechanism (RCM) disrupted the global structure into the local regions and finds the discrimination area by the feature extraction module. Nevertheless, adversarial learning is introduced into the DL part, which prevents the noise emerged by overfitting RCM from negatively affecting network learning. In addition, CL induces the feature extraction module to learn the semantic correlative among local regions to reconstitute the original image. In the inference process, test samples have been sent into the trained feature extraction network and obtained the classification results.

In this paper, ECADCL is designed to learn the local feature space based on the disruption of the global structure. The overall architecture structure of ECADCL is illustrated in Figure 4, and in the following subsections, we will present our proposed tool wear monitoring method based on machined surface images in detail.

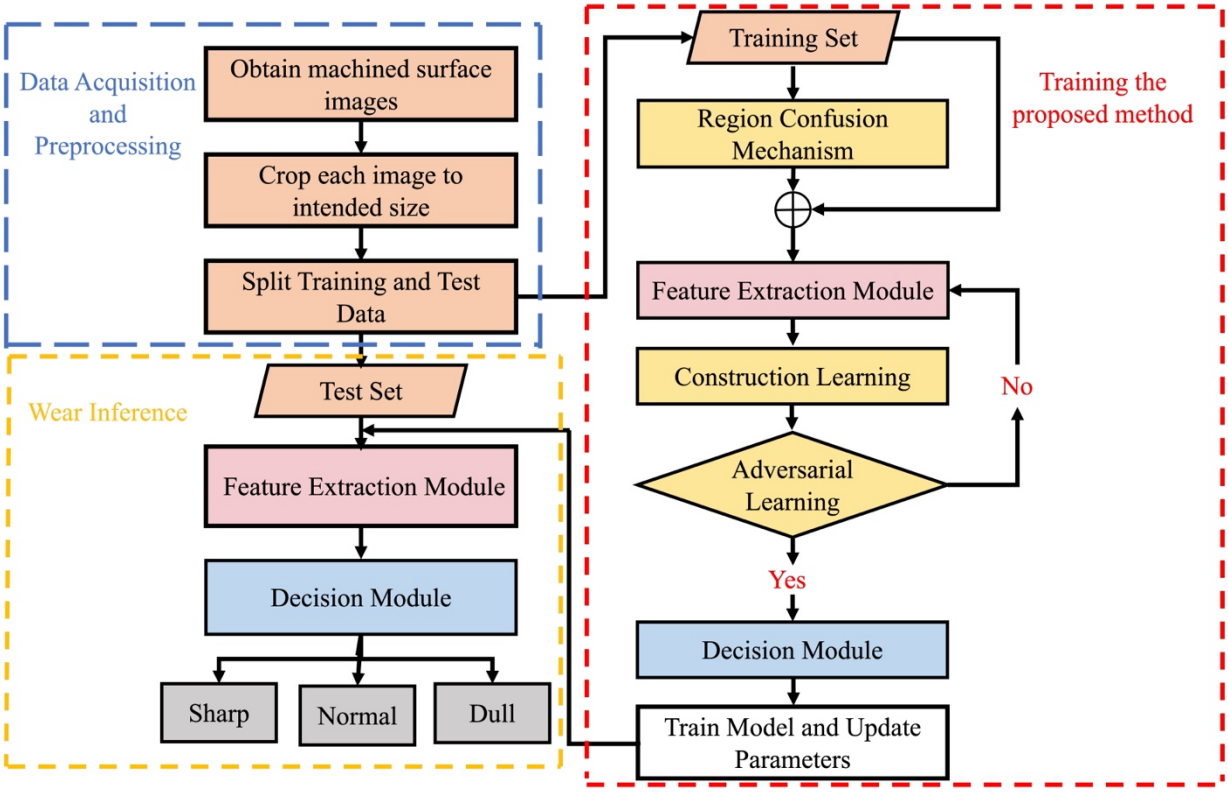


Figure 3. Flow chart of the proposed method for tool wear monitoring.

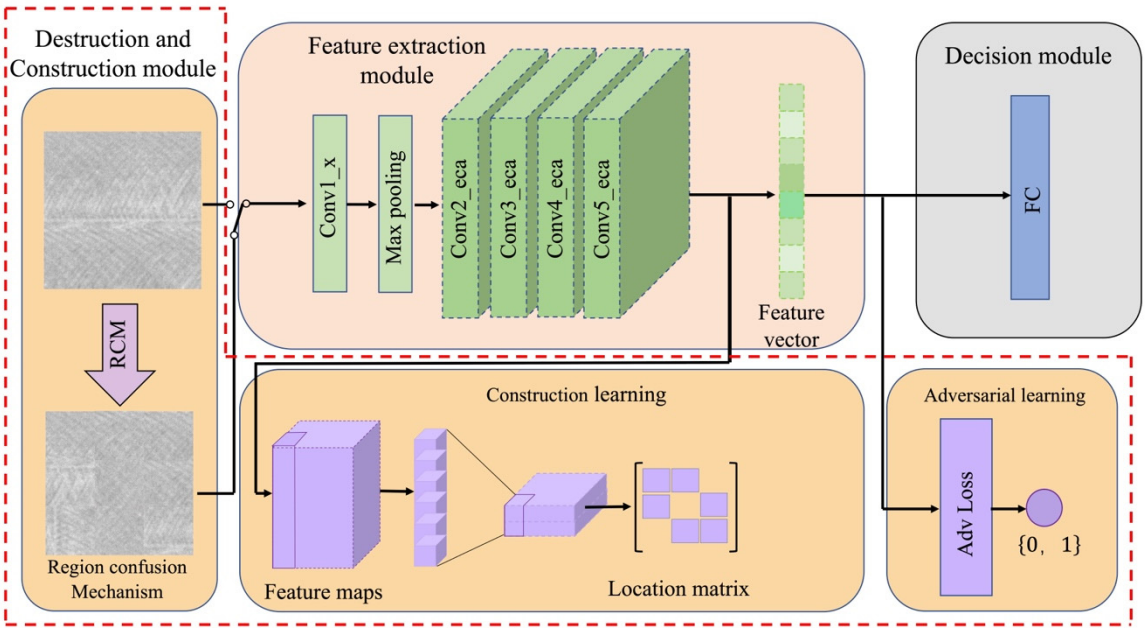


Figure 4. The architecture of ECADCL, which includes three modules: (1) feature extraction module; (2) destruction and construction module; (3) Decision module.

3.1. Feature Extraction Module

Our suggested framework is a typical CNN, which can be seen in Figure 3. The creation of a network structure is a significant challenge in practical applications. However, transfer learning has been found to be a well-performing way to overcome this issue. In this study, the ECADCL architecture is developed, relying on three types of famous networks in transfer learning, namely, AlexNet [42], VGG-16 [43], and ResNet-18 [44], which exhibit

excellent performance for image recognition. Comparing the results of the above three famous networks, the ResNet18 is selected as the basic framework, and the details of the proposed model architecture are listed in Table 2.

Table 2. Architectural details of the proposed model based on ResNet-18.

Name	Parameters	Output Size
Input layer	Image size: 256 × 256 × 3	256 × 256 × 3
Conv1_x	Filter: [7 × 7, 64], stride: [2 2]	128 × 128 × 6
Conv2_eca	3 × 3 max pooling, stride: [2 2] Filter: $\begin{bmatrix} 3\times 3, & 64 \\ 3\times 3, & 64 \\ \text{ECA} & 64 \end{bmatrix} \times 2$	64 × 64 × 64
Conv3_eca	Filter: $\begin{bmatrix} 3\times 3, & 64 \\ 3\times 3, & 128 \\ \text{ECA} & 128 \end{bmatrix} \times 2$	32 × 32 × 128
Conv4_eca	Filter: $\begin{bmatrix} 3\times 3, & 128 \\ 3\times 3, & 256 \\ \text{ECA} & 256 \end{bmatrix} \times 2$	16 × 16 × 256
Conv5_eca	Filter: $\begin{bmatrix} 3\times 3, & 256 \\ 3\times 3, & 256 \\ \text{ECA} & 512 \end{bmatrix} \times 2$	8 × 8 × 512

As illustrated in Table 2, the feature extraction module architecture, mainly constituted by one input layer and five convolutional blocks, has the same size as the input image, 256 × 256 × 3. There is a max pooling layer, 8 ECA layers, and 17 convolutional layers in the five convolutional blocks. Following the first convolutional block, the max pooling layer is responsible for replacing the output for the appropriate places with the overall features of the nearby domains in order to complete the downsampling procedure. In this study, the rectified linear unit (ReLU) activation function is used for nonlinearity. Note that the original image $\mathcal{X}$, its destroyed version $\phi(\mathcal{X})$, and its ground-truth one-vs.-all labelling $\mathcal{Y}$ are combined for training. Therefore, the extracted features may be described by the equation:

$$F_t = \left\langle C_n^m(\mathcal{X}) C_n^m(\phi(\mathcal{X})) \right\rangle = \forall(I_i \,|\theta_m) \in \mathbb{R}^{W\times H\times N} \quad (1)$$

where F_t indicates the learned features, $C_n^m(\mathcal{X})$ represents the mth feature of the original image in the nth channels, $\forall$ denotes the feature extraction procedure, θ represents all the learnable parameters, and the three dimensions indicate the width, height, and number of channels, respectively, of the learned features.

3.2. Attention-Based Network

The underlying attention mechanisms are inspired by the human visual system, which tends to process the most critical information in a scene rather than attempting to analyze the whole scene simultaneously. Accordingly, an attention mechanism can be introduced into a CNN to concentrate on more representative sections while suppressing less significant information. For this purpose, the ECA module is introduced into our feature extraction module. Referring to Figure 5, EAC is a local cross-channel interaction strategy that can be efficiently implemented without dimensionality reduction to adaptively determine the coverage of local cross-channel interactions.

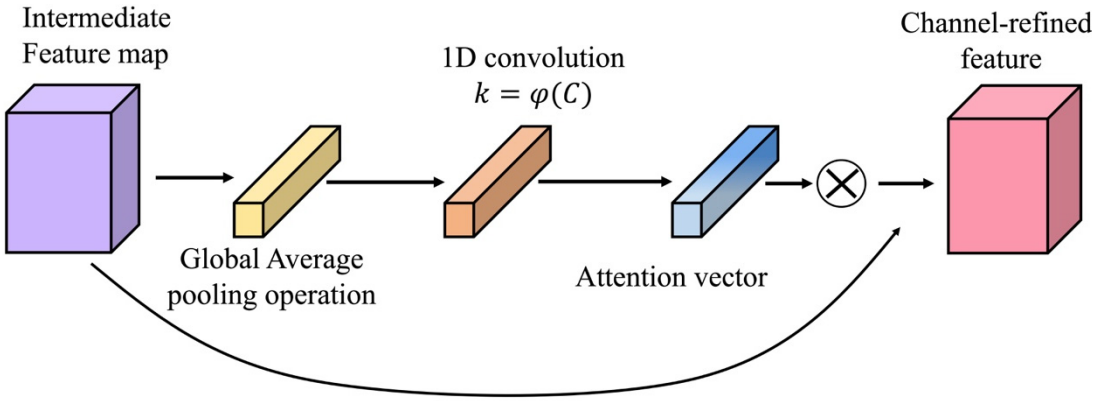


Figure 5. Structure of the ECA module.

In the ECA block, the intermediate feature map F_t is subjected to the global average pooling (GAP) operation and then generates an aggregated feature vector: V_{agg}. This vector is then input into a fast 1D convolution of size k, which produces the attention vector A_{eca}. Finally, to acquire the channel attention vector V_c, the attention vector is passed via the sigmoid activation function. Thus, V_c can be generated as follows:

$$V_c = \sigma(C1D_k(G(F_t))) = \sigma\left(C1D_k\left(V_{agg}\right)\right) = \sigma(A_{eca}) \quad (2)$$

where $G(\cdot)$ denotes the GAP operator, $C1D_k(\cdot)$ represents a 1D convolutional layer and its kernel size k, and $\sigma(\cdot)$ is the sigmoid function.

In the ECA block, determining the value of k is another critical issue. Considering a similar principle to group convolutions, different channel dimensions C correspond to different convolution ranges k in feature maps, giving rise to a mapping relationship between k and C. This relationship is characterized by a nonlinear function mapping from kernel size k to channel dimension C, which can be expressed adaptively as follows:

$$k = \varphi(C) = \left|\frac{\log_2(C)}{\gamma} + \frac{a}{\gamma}\right|_{odd} \quad (3)$$

where $|t|_{odd}$ is the odd integer that is closest to t. As verified in [35], we set γ and a to 2 and 1, respectively. $\varphi(\cdot)$ represents a nonlinear mapping that relates feature maps with various dimensions of channels and varying interaction ranges, driving the model to adaptively learn the interdependencies between feature channels.

3.3. Destruction-Construction Module

As mentioned above, machined surface images are visually similar in terms of their global information. Hence, the proposed algorithm needs to learn to discriminate local details within the same global structure. The destruction-construction strategy is introduced to improve the ECACNN architectures to address this issue. The destruction-construction module consists of three branches: the region confusion mechanism (RCM), adversarial learning, and construction learning. RCM is the key part of the destruction-construction module since it destroys the entire structure by causing localized disruption. To avoid the noise caused by the RCM operation, adversarial learning is recommended during the destruction learning process. Construction learning is proposed for reorganizing previously learned local information in accordance with the semantic relevance among areas. For the input, consisting of the original $\mathcal{X}$, and after RCM obtained destroyed images $\phi(\mathcal{X})$, they are sent into the feature extraction module to obtain the feature map, which is sent to the construction module, while the adversarial network receives the feature map after average pooling.

For RCM, it is suggested that the spatial distribution of the whole structure be disrupted and allow the local areas to move within a specific range. In the confusion method,

any original image $\mathcal{X}$ is divided into N × N subregions, each denoted by $\mathcal{X}_{i,j}$, which corresponds to the horizontal coordinate i and vertical coordinates j. The subregions are rearranged as a destroyed image $\phi(\mathcal{X})$, which generates a new coordinate:

$$\sigma(i,j) = \left(\sigma_j^{row}(i), \sigma_i^{col}(j)\right) \tag{4}$$

where $\sigma(i,j)$ represents the subregions at locating (i,j) in the original image $\mathcal{X}$, $\sigma_j^{row}(i)$ represents a new coordinate σ_j^{row} of the areas in the jth row of $\mathcal{X}$ by a random vector q_j of size N, the size of the ith element $q_{j,i} = i + r$, with $r \sim U(-k, k)$ being a uniformly distributed random variable within the stretch of $[-k, k]$, and k as a tunable parameter ($1 < k < N$) defines the neighborhood range. $\sigma_i^{col}(j)$ similarly represents a new coordinate σ_i^{col} of the area in the ith column of $\mathcal{X}$.

For adversarial learning, a discriminator is designed as a new stream to judge whether the input image $\mathcal{X}$ has been destroyed or not. It prevents the RCM-induced noise patterns from shuffling the local regions into the feature map for classification tasks, which can be written as follows:

$$D(\mathcal{X}, \theta_{adv}) = \text{softmax}\left(\theta_{adv} F\left(\mathcal{X}, \theta_i^{[1,m]}\right)\right) \tag{5}$$

where $F\left(\mathcal{X}, \theta_i^{[1,m]}\right)$ is the feature vector extracted from the outputs of the mth layer in the feature extraction module, $\theta_i^{[1,m]}$ represents the learnable parameters from the 1st layer to the mth layers in the feature extraction module, and $\theta_{adv} \in \mathcal{R}^{d \times 2}$ is a linear mapping.

For construction learning, a region alignment network is designed to calculate the positional precision of different regions in the images and simulate the correlations between local regions in an end-to-end fashion. This can be expressed as follows:

$$M(\mathcal{X}) = h\left(F\left(\mathcal{X}, \theta_i^{[1,n]}\right), \theta_{loc}\right) \tag{6}$$

where $M(\mathcal{X})$ represents the two channels of row and column coordinates, $h(\cdot)$ represents the suggested region alignment network, $F\left(\mathcal{X}, \theta_i^{[1,n]}\right)$ is the feature vector extracted from the outputs of the nth layer in the feature extraction module, $\theta_i^{[1,m]}$ represents the learnable parameters from the 1st layer to the nth layers in the feature extraction module, and θ_{loc} represents the learnable parameters in the region alignment network.

The final decision module, relying on the typical classified design of the network, includes a fully connected layer to predict the output results.

4. Case Validation

4.1. Dataset

In this study, we acquired a large number of images from the image acquisition system to train the network. Since the quality of the workpiece images is affected by an illuminated environment, the raw images were center-cropped into RGB images with dimensions of 2560 × 2560 pixels to reduce the effect on the brightness uniformity for further analysis. To facilitate the experiment without model overfitting, a cropped RGB image was uniformly cropped to 100 sub-images of 256 × 256 pixels. Eventually, the raw dataset is made from a total of 18,000 images of machined surface images involving four cutting tools. The raw dataset was divided into a training set and a test set, including 12,600 training images and 5400 test images. The distribution of dataset samples is labelled in Table 3.

Table 3. The distribution of the dataset.

Dataset	Training Set	Test Set
Sharp	4200	1800
Normal	4200	1800
Dull	4200	1800

4.2. Implementation Details

The laboratory environment: Python 3.9, PyTorch 1.10.0, and torchvision 0.11.1. For all methods in the experiments, 200 training epochs were performed; the initial learning rate was 0.0001, and it decayed by a factor of 10 at 40, 80, and 120 epochs; and the training set was divided into 64 batches. The number of workers in PyTorch was set to eight for both training and validation. To train the proposed method, in the RCM, the number of regions N must be divisible by the size of an input image, and was set to four in this study. Then, the proposed feature extraction module, ECACNN, was used to extract the feature map from the input, which was sent to the construction module while it was sent to the adversarial network after average pooling. Finally, the feature vector is fed into the decision module to predict the outcome of the classification category. For the testing period, the destruction-construction module is disabled. At the same time, the original images as input are directly sent to the feature extraction module and decision module for the prediction of the results.

4.3. Architecture Comparison with ECA

In the experiment, three types of well-known networks, namely AlexNet, VGG-16, and ResNet-18, were used as the base frameworks to build and compare the networks. For fairness, the ECA module was added after each convolutional block. Table 4 lists the experimental results, which demonstrate that the accuracy rate obtained when using the VGG-16 architecture to construct the ECACNN is higher than that obtained when using AlexNet and ResNet-18. On this image classification task, the classification accuracy of the ECACNN based on the AlexNet architecture reached 87.56%, while the classification accuracy of the ECACNNs based on the ResNet-18 and VGG-16 architectures obtained 92.32% and 94.04%, respectively. This result emerged that the depth of architectures from AlexNet to VGG-16 was increased, enabling the extraction of high-level and complex abstractions' information from raw data. It should be noted that ResNet-18 and VGG-16 have a similar size in depth, but ResNet-18 has a substantially lower level of complexity than VGG-16, which is a significant advantage. Specifically, the number of parameters of ResNet-18 (11.18 million) is equivalent to only 13.0% of the parameters in VGG-16 (86.03 million). Therefore, in this study, ResNet-18 was chosen as the backbone of the ECACNN.

Table 4. The recognition accuracy of ECACNN based on different architectures.

Network Architecture	Accuracy (%)	Parameters (Million)
AlexNet	87.56	33.64
VGG-16	94.09	86.03
ResNet-18	92.32	11.18

4.4. Results of Comparing ECADCL with Other Methods

To evaluate the effectiveness of the proposed model, we compared with two traditional models and other classical convolutional neural models for tool wear monitoring based on machined surface images. Traditional methods included principal component analysis (PCA) and genetic algorithm-support vector machine (GA-SVM) traditional handcraft method, GLCM features and GA-SVM traditional handcraft method. Traditional machine learning techniques are constructed by manually extracting features and machine learning models. For feature extracting, the GLCM features extracted as the handcraft methods are specified in Table 5. The PCA method for feature extraction set the accumulative variance percentage at 0.95 as default. At the same time, the machine learning models by using GA-SVM when analyzing extraction features for monitoring the three different levels of cutting tool. The popular deep learning methods of AlexNet, Vgg-16, and ResNet-18 were directly applied for classification. The evaluation results of different methods are presented in Table 6, where it can be seen that the deep learning method outperforms traditional handcraft approaches in classifying tool wear state.

Table 5. The handcrafted GLCM features list.

Feature	Formula
Contrast	$\sum_{i,j=0}^{levels-1} P_{i,j}(i-j)^2$
Dissimilarity	$\sum_{i,j=0}^{levels-1} P_{i,j}\lvert i-j\rvert$
Homogeneity	$\sum_{i,j=0}^{levels-1} \frac{P_{i,j}}{1+(i-j)^2}$
ASM	$\sum_{i,j=0}^{levels-1} P_{i,j}^2$
Correlation	$\sum_{i,j=0}^{levels-1} P_{i,j}\left[\frac{(i-\mu_i)(j-\mu_j)}{\sigma_i \sigma_j}\right]$

Table 6. Performance comparison.

Method		Accuracy (%)
Traditional	PCA + GA-SVM	59.56
	GLCM features + GA-SVM	79.80
Deep learning	AlexNet	88.68
	Vgg-16	86.68
	ResNet-18	90.29
	ECACNN	92.32
	Ours, ECADCL	99.96

The superiority of the DCL approach for tool wear monitoring is investigated in this section. A second experiment was carried out to determine the precise impact of the proposed model. We designed additional models by using the ECACNN as the backbone network and introducing either the destruction stream or the construction stream individually. The experimental results are shown in Table 6. The accuracy rate of the proposed ECADCL method is 99.96%, representing a significant boost in performance. In our proposed method, destruction learning (DL), which enables discrimination from noisy, detailed, and global features, is verified to help improve the performance of classification. Likewise, the shapes and compositions of the formations of objects modelled by construction learning (CL) can further enhance the performance of the classification model.

4.5. Ablation Experiments

In order to prove the feasibility of the proposed model, an ablation experiment is utilized for validation of different parts in ECADCL, including building four modules and comparing their recognition accuracy. The SE module, as channel attention, was compared with ECA in the feature extraction module. The experimental results of each module are recorded in Table 7.

Table 7. The ablation experimental results.

Method	Accuracy (%)
ResNet-18 (baseline)	90.29
+SE	91.61
+ECA	92.32
+ECA+RCM	99.36
+ECA+RCM+ Adversarial Learning (DL)	99.91
+ECA+RCM+ Construction Learning (CL)	99.78
Ours, ECADCL	99.96

As shown in Table 7, the ablation experiment demonstrates that the performance improvement caused by RCM prove that learning the detailed visual pattern from a "shuffled" visual feature space is beneficial to recognition texture. The ablation experiment also demonstrates that the feature learning ability of the ECA module is better than the SE

in the feature extracting module. It can be observed that the ECA's destruction-construction-based image classification is capable of dealing with tool condition monitoring tasks.

4.6. Discussion

The proposed ECADCL algorithm exhibits excellent performance in tool wear monitoring, and four metrics were used to evaluate the efficacy of the proposed method: confusion matrix, precision, recall, and F-score. The confusion matrix describes a comprehensive overview of the method performance, which is shown in Figure 6. The x- and y-axes reflect the predicted and actual labels, respectively, while the diagonal elements represent the accuracy of each particular state.

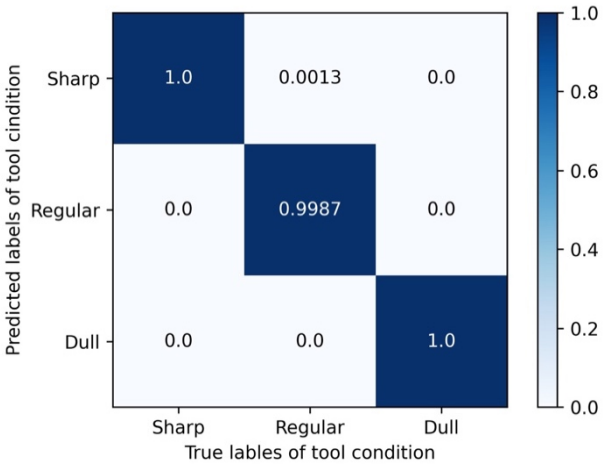


Figure 6. Confusion matrix of the proposed method.

We explore the features learned by ECADCL by visualizing the last convolutional layers of Conv1_x, Conv2_eca, Conv3_eca, and Conv5_eca in Figure 6. In Figure 7a, it can be seen that filters are primarily concerned with extracting texture characteristics, which enables the proposed method to extract more complicated features from deeper layers. From Figure 7b–d, it is noted that more specific information is provided from the superficial layers to the deeper layers.

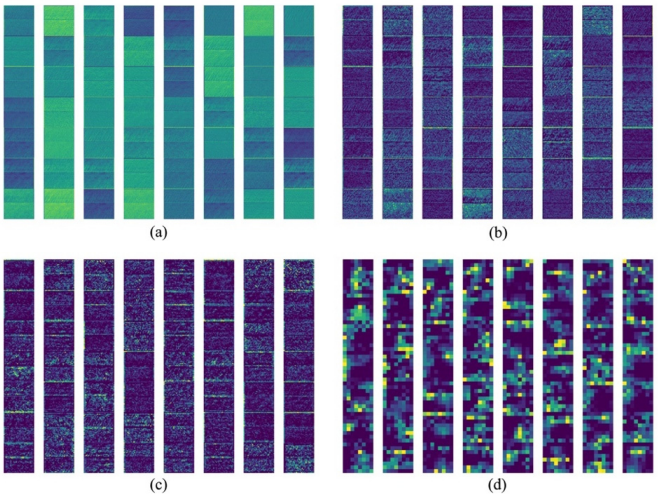


Figure 7. Visualization of the feature maps from (**a**) the last convolutional layer of Conv1_x. (**b**) Last convolutional layer of Conv2_eca. (**c**) Last convolutional layer of Conv3_eca. (**d**) Last convolutional layer of Conv5_eca.

In addition, we focus on how the running time changes with respect to the model. Deep learning algorithms have the advantage of being super-fast at testing. In order to

highlight the superiority of the proposed method in the field of monitoring tool wear, the running time was compared to ResNet-18 and ECACNN on the 5400-sample dataset. The running time of ECADCL, ECACNN, and ResNet-18 are 78 s, 76 s, and 73 s, respectively. Experiments results demonstrated that, compared to other classification algorithms in the control group, the proposed method greatly improved accuracy without losing computation time.

5. Conclusions

In this article, we propose a new tool wear monitoring framework with an emphasis on fine-grained categorization and deep learning technology. AlexNet, VGG-16, and ResNet-18, three types of classical deep learning architectures, were researched and compared as a base network to develop an effective network. Introducing the ECA module into the base network enables the network to focus on more representative parts of the input image while suppressing less critical information. The destruction-construction module is also employed to discriminate local features, because of the similar global structures in the surface texture images. In experiments, our proposed model proves more effective than other methods for monitoring the level of tool wear.

However, to train the proposed model for machined surface-based tool wear state monitoring, a deep learning model requires a large amount of training data to achieve outstanding monitoring performance. As a result, our future work will focus on suitable unsupervised models to overcome this challenge. Meanwhile, in the image acquisition process, since the contamination of chips and crude oils generate machined surface pollution, it poses a great analytical challenge for an intelligent monitoring tool wear system. Furthermore, variable factors, such as various tools and cutting parameters, will be explored to further enhance the suggested method's universality.

Author Contributions: Conceptualization, J.Y.; methodology, J.Y. and J.D.; software, J.Y.; validation, J.Y., J.D. and T.L.; formal analysis, J.Y.; investigation, J.Y., J.D. and C.H; data curation, J.Y. and J.L.; writing—original draft preparation, J.Y.; writing—review and editing, J.D., T.L., C.H., J.L. and T.S.; supervision, T.S.; visualization, J.Y.; project administration, T.S.; funding acquisition, T.S. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the Key-Area Research and Development Program of Guangdong Province (No. 2020B090927002), the National Natural Science Foundation of China (No. 52205103), the National Key Research and Development Program of China (No. 2020YFB1709800).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Not applicable.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Haraguchi, N.; Cheng, C.F.C.; Smeets, E. The Importance of Manufacturing in Economic Development: Has This Changed? *World Dev.* **2017**, *93*, 293–315. [CrossRef]
2. Luan, X.; Zhang, S.; Li, J.; Mendis, G.; Zhao, F.; Sutherland, J.W. Trade-off Analysis of Tool Wear, Machining Quality and Energy Efficiency of Alloy Cast Iron Milling Process. *Procedia Manuf.* **2018**, *26*, 383–393. [CrossRef]
3. Liu, C.; Wang, G.F.; Li, Z.M. Incremental Learning for Online Tool Condition Monitoring Using Ellipsoid ARTMAP Network Model. *Appl. Soft Comput.* **2015**, *35*, 186–198. [CrossRef]
4. Zhou, Y.; Sun, B.; Sun, W. A Tool Condition Monitoring Method Based on Two-Layer Angle Kernel Extreme Learning Machine and Binary Differential Evolution for Milling. *Measurement* **2020**, *166*, 108186. [CrossRef]
5. Mohanraj, T.; Shankar, S.; Rajasekar, R.; Sakthivel, N.R.; Pramanik, A. Tool Condition Monitoring Techniques in Milling Process—A Review. *J. Mater. Res. Technol.* **2020**, *9*, 1032–1042. [CrossRef]
6. Fernández-Robles, L.; Sánchez-González, L.; Díez-González, J.; Castejón-Limas, M.; Pérez, H. Use of Image Processing to Monitor Tool Wear in Micro Milling. *Neurocomputing* **2021**, *452*, 333–340. [CrossRef]
7. Peng, R.; Pang, H.; Jiang, H.; Hu, Y. Study of Tool Wear Monitoring Using Machine Vision. *Autom. Control Comput. Sci.* **2020**, *54*, 259–270. [CrossRef]

8. Azmi, A.I. Monitoring of Tool Wear Using Measured Machining Forces and Neuro-Fuzzy Modelling Approaches during Machining of GFRP Composites. *Adv. Eng. Softw.* **2015**, *82*, 53–64. [CrossRef]
9. Cao, X.C.; Chen, B.Q.; Yao, B.; He, W.P. Combining Translation-Invariant Wavelet Frames and Convolutional Neural Network for Intelligent Tool Wear State Identification. *Comput. Ind.* **2019**, *106*, 71–84. [CrossRef]
10. Li, T.; Shi, T.; Tang, Z.; Liao, G.; Duan, J.; Han, J.; He, Z. Real-Time Tool Wear Monitoring Using Thin-Film Thermocouple. *J. Mater. Process. Technol.* **2021**, *288*, 116901. [CrossRef]
11. Yen, C.L.; Lu, M.C.; Chen, J.L. Applying the Self-Organization Feature Map (SOM) Algorithm to AE-Based Tool Wear Monitoring in Micro-Cutting. *Mech. Syst. Signal Process.* **2013**, *34*, 353–366. [CrossRef]
12. Bhat, N.N.; Dutta, S.; Vashisth, T.; Pal, S.; Pal, S.K.; Sen, R. Tool Condition Monitoring by SVM Classification of Machined Surface Images in Turning. *Int. J. Adv. Manuf. Technol.* **2015**, *83*, 1487–1502. [CrossRef]
13. Dutta, S.; Pal, S.K.; Sen, R. On-Machine Tool Prediction of Flank Wear from Machined Surface Images Using Texture Analyses and Support Vector Regression. *Precis. Eng.* **2016**, *43*, 34–42. [CrossRef]
14. Dutta, S.; Pal, S.K.; Sen, R. Tool Condition Monitoring in Turning by Applying Machine Vision. *J. Manuf. Sci. Eng. Trans. ASME* **2016**, *138*, 051008. [CrossRef]
15. Kassim, A.A.; Mannan, M.A.; Jing, M. Machine Tool Condition Monitoring Using Workpiece Surface Texture Analysis. *Mach. Vis. Appl.* **2000**, *11*, 257–263. [CrossRef]
16. Riego, V.; Castejón-Limas, M.; Sánchez-González, L.; Fernández-Robles, L.; Perez, H.; Diez-Gonzalez, J.; Guerrero-Higueras, Á.M. Strong Classification System for Wear Identification on Milling Processes Using Computer Vision and Ensemble Learning. *Neurocomputing* **2021**, *456*, 678–684. [CrossRef]
17. Li, L.; An, Q. An In-Depth Study of Tool Wear Monitoring Technique Based on Image Segmentation and Texture Analysis. *Measurement* **2016**, *79*, 44–52. [CrossRef]
18. Liu, Y.; Guo, L.; Gao, H.; You, Z.; Ye, Y.; Zhang, B. Machine Vision Based Condition Monitoring and Fault Diagnosis of Machine Tools Using Information from Machined Surface Texture: A Review. *Mech. Syst. Signal Process.* **2022**, *164*, 108068. [CrossRef]
19. Lecun, Y.; Bengio, Y.; Hinton, G. Deep Learning. *Nature* **2015**, *521*, 436–444. [CrossRef]
20. Wu, Y.; Guo, H.; Chakraborty, C.; Khosravi, M.; Berretti, S.; Wan, S. Edge Computing Driven Low-Light Image Dynamic Enhancement for Object Detection. *IEEE Trans. Netw. Sci. Eng.* **2022**. [CrossRef]
21. Liu, Z.T.; Wu, B.H.; Li, D.Y.; Xiao, P.; Mao, J.W. Speech Emotion Recognition Based on Selective Interpolation Synthetic Minority over-Sampling Technique in Small Sample Environment. *Sensors* **2020**, *20*, 2297. [CrossRef] [PubMed]
22. Park, S.; Park, S.; Jeong, H.; Yun, I.; So, J. Scenario-Mining for Level 4 Automated Vehicle Safety Assessment from Real Accident Situations in Urban Areas Using a Natural Language Process. *Sensors* **2021**, *21*, 6929. [CrossRef] [PubMed]
23. He, Z.; Shi, T.; Xuan, J.; Li, T. Research on Tool Wear Prediction Based on Temperature Signals and Deep Learning. *Wear* **2021**, *478–479*, 203902. [CrossRef]
24. Bazi, R.; Benkedjouh, T.; Habbouche, H.; Rechak, S.; Zerhouni, N. A Hybrid CNN-BiLSTM Approach-Based Variational Mode Decomposition for Tool Wear Monitoring. *Int. J. Adv. Manuf. Technol.* **2022**, *119*, 3803–3817. [CrossRef]
25. 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]
26. He, Z.; Shi, T.; Xuan, J. Milling Tool Wear Prediction Using Multi-Sensor Feature Fusion Based on Stacked Sparse Autoencoders. *Measurement* **2022**, *190*, 110719. [CrossRef]
27. Brili, N.; Ficko, M.; Klančnik, S. Automatic Identification of Tool Wear Based on Thermography and a Convolutional Neural Network during the Turning Process. *Sensors* **2021**, *21*, 1917. [CrossRef] [PubMed]
28. Duan, J.; Zhang, X.; Shi, T. A Hybrid Attention-Based Paralleled Deep Learning Model for Tool Wear Prediction. *Expert Syst. Appl.* **2023**, *211*, 118548. [CrossRef]
29. Kumar, M.P.; Dutta, S.; Murmu, N.C. Tool Wear Classification Based on Machined Surface Images Using Convolution Neural Networks. *Sadhana-Acad. Proc. Eng. Sci.* **2021**, *46*, 130. [CrossRef]
30. Chen, Y.; Zhang, X.; Chen, Z.; Song, M.; Wang, J. Fine-Grained Classification of Fly Species in the Natural Environment Based on Deep Convolutional Neural Network. *Comput. Biol. Med.* **2021**, *135*, 104655. [CrossRef] [PubMed]
31. Wang, W.; Cui, Y.; Li, G.; Jiang, C.; Deng, S. A Self-Attention-Based Destruction and Construction Learning Fine-Grained Image Classification Method for Retail Product Recognition. *Neural Comput. Appl.* **2020**, *32*, 14613–14622. [CrossRef]
32. Yu, Y.; Xu, L.; Jia, W.; Zhu, W.; Fu, Y.; Lu, Q. CAM: A Fine-Grained Vehicle Model Recognition Method Based on Visual Attention Model. *Image Vis. Comput.* **2020**, *104*, 104027. [CrossRef]
33. Song, A.P.; Hu, Q.; Ding, X.H.; Di, X.Y.; Song, Z.H. Similar Face Recognition Using the IE-CNN Model. *IEEE Access* **2020**, *8*, 45244–45253. [CrossRef]
34. Lin, Z.; Jia, J.; Gao, W.; Huang, F. Fine-Grained Visual Categorization of Butterfly Specimens at Sub-Species Level via a Convolutional Neural Network with Skip-Connections. *Neurocomputing* **2020**, *384*, 295–313. [CrossRef]
35. Kim, H.; Kim, H.J.; Park, J.; Ryu, J.K.; Kim, S.C. Recognition of Fine-Grained Walking Patterns Using a Smartwatch with Deep Attentive Neural Networks. *Sensors* **2021**, *21*, 6393. [CrossRef]
36. Chen, Y.; Bai, Y.; Zhang, W.; Mei, T. Destruction and Construction Learning for Fine-Grained Image Recognition. In Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, Long Beach, CA, USA, 15–20 June 2019; pp. 5152–5161. [CrossRef]

37. Du, X.; Wang, J.; Sun, W.; Liu, H.; Liu, Z.; Jia, W.; Zeng, Y.; Liu, R.; Liu, X. A Novel Approach to Tool Condition Monitoring Based on Multi-Sensor Data Fusion Imaging and an Attention Mechanism. *Meas. Sci. Technol.* **2021**, *32*, 055601. [CrossRef]
38. Yang, G.; He, Y.; Yang, Y.; Xu, B. Fine-Grained Image Classification for Crop Disease Based on Attention Mechanism. *Front. Plant Sci.* **2020**, *11*, 600854. [CrossRef]
39. Li, H.; Shao, X.; Zhang, C.; Qian, X. Automated Assessment of Parkinsonian Finger-Tapping Tests through a Vision-Based Fine-Grained Classification Model. *Neurocomputing* **2021**, *441*, 260–271. [CrossRef]
40. Hu, J.; Shen, L.; Albanie, S.; Sun, G.; Wu, E. Squeeze-and-Excitation Networks. *IEEE Trans. Pattern Anal. Mach. Intell.* **2020**, *42*, 2011–2023. [CrossRef]
41. Wang, Q.; Wu, B.; Zhu, P.; Li, P.; Zuo, W.; Hu, Q. ECA-Net: Efficient Channel Attention for Deep Convolutional Neural Networks. In Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, Seattle, WA, USA, 14–19 June 2020; pp. 11531–11539. [CrossRef]
42. Krizhevsky, A.; Sutskever, I.; Hinton, G. ImageNet Classification with Deep Convolutional Neural Networks. In Proceedings of the NIPS 2012: 26th Annual Conference on Neural Information Processing Systems 2012, Lake Tahoe, NV, USA, 3–6 December 2012; pp. 1097–1105.
43. Simonyan, K.; Zisserman, A. Very Deep Convolutional Networks for Large-Scale Image Recognition. In Proceedings of the 3rd International Conference on Learning Representations, ICLR 2015—Conference Track Proceedings, San Diego, CA, USA, 7–9 May 2015. [CrossRef]
44. He, K.; Zhang, X.; Ren, S.; Sun, J. Deep Residual Learning for Image Recognition. In Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 27–30 June 2016; pp. 770–778. [CrossRef]