

Article

High-Frequency Workpiece Image Recognition Model Integrating Multi-Level Network Structure

Yang Ou ¹ , Chenglong Sun ², Rong Yuan ¹ and Jianqiao Luo ^{3,*}

¹ School of Mechanical Engineering, Chengdu University, Chengdu 610106, China; ouyang@my.swjtu.edu.cn (Y.O.); yuanrong27@126.com (R.Y.)

² School of Mechanical Engineering, Southwest Jiaotong University, Chengdu 610031, China; sunclwlkq@163.com

³ College of Electrical and Information Engineering, Hunan University, Changsha 410082, China

* Correspondence: jqluo@hnu.edu.cn; Tel.: +86-182-1567-5028

Abstract: High-frequency workpieces have the characteristics of complex intra-class textures and small differences between classes, leading to the problem of low recognition rates when existing models are applied to the recognition of high-frequency workpiece images. We propose in this paper a novel high-frequency workpiece image recognition model that uses EfficientNet-B1 as the basic network and integrates multi-level network structures, designated as ML-EfficientNet-B1. Specifically, a lightweight mixed attention module is first introduced to extract global workpiece image features with strong illumination robustness, and the global recognition results are obtained through the backbone network. Then, the weakly supervised area detection module is used to locate the locally important areas of the workpiece and is introduced into the branch network to obtain local recognition results. Finally, the global and local recognition results are combined in the branch fusion module to achieve the final recognition of high-frequency workpiece images. Experimental results show that compared with various image recognition models, the proposed ML-EfficientNet-B1 model has stronger adaptability to illumination changes, significantly improves the performance of high-frequency workpiece recognition, and the recognition accuracy reaches 98.3%.

Keywords: network structure; hybrid attention; image recognition; deep learning



Citation: Ou, Y.; Sun, C.; Yuan, R.; Luo, J. High-Frequency Workpiece Image Recognition Model Integrating Multi-Level Network Structure.

Sensors **2024**, *24*, 1982. <https://doi.org/10.3390/s24061982>

Academic Editor: Sheryl Berlin Brahnam

Received: 31 January 2024

Revised: 5 March 2024

Accepted: 19 March 2024

Published: 20 March 2024



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

With the advancement of science and technology, the global manufacturing industry is developing rapidly towards high quality, high efficiency, and high intelligence [1–3]. High-frequency workpieces are one of the most important components in aerospace equipment. The quality, timeliness, and intelligence of their processing are all important factors affecting the development of the aerospace industry. During the processing of the workpiece, the workpiece is bound to the pallet with an RFID tag, and the processing content to be completed in the process is automatically loaded through scanning by the RFID tag reading and writing device. However, during the heat treatment stage of the workpiece, the separation of the workpiece from the pallet causes the failure of the original radio frequency chip of the workpiece, which affects the automation and intelligence of subsequent processing. Therefore, image recognition technology is used to identify the workpiece image after heat treatment and the workpiece image before heat treatment and then associate the processing content corresponding to the image number again and attach a new RFID tag to the high-frequency workpiece after heat treatment. The new RFID tag is the same as the tag before heat treatment. Radiofrequency chips with the same content are used to complete the intelligent process of subsequent finishing and other processing procedures. In order to improve the intelligence level of high-frequency workpiece processing, image recognition technology can be introduced into the processing process. The image recognition results are associated with the drawing number of the currently processed high-frequency workpiece,

and the next step of processing can be automatically loaded through the manufacturing execution system (MES). At present, image recognition of high-frequency workpieces faces the following challenges: (1) The same types of workpieces have complex internal textures. (2) Different types of workpieces have small differences in characteristics. (3) The quality of the workpiece image is greatly affected by changes in acquisition posture and lighting.

Image recognition technology has been widely used in different fields, and many researchers have proposed a variety of recognition algorithms [4–19]. In [13], the author built a five-layer memristor-based CNN to perform MNIST10 image recognition and achieved a high accuracy of more than 96%. In [15], the author proposed a fastener classification model, which can divide fasteners into four types, including normal, partially worn, missing, and covered. In the industrial field, researchers have proposed some effective algorithms for image recognition of mechanical parts. It can be divided into traditional algorithms based on mathematical models and deep learning algorithms based on convolutional neural networks. In the first category, Xu et al. proposed a template matching algorithm RTMM using ring feature templates to solve the problem of low accuracy of traditional matching methods caused by the presence of different sides of parts in the image [20]. Yin et al. proposed a fast positioning and recognition algorithm based on equal-area ring segmentation to solve the problem of complex textures and high similarity after precision machining of high-frequency components, which makes it difficult to distinguish [21]. Wang et al. proposed a method of part sorting based on fast template matching, which accelerates the process of part recognition by improving the template matching method. Furthermore, the efficiency of the part sorting system based on machine vision is improved [22]. In the second category, Yang et al. proposed a joint loss-supervised deep learning recognition algorithm. This algorithm first builds an image feature vector encoding model based on a convolutional neural network, uses angle margin loss to replace SoftMax loss to reduce the distance between features within the workpiece class, and then introduces isolation loss to increase the distance between features of heterogeneous workpieces [23]. Zhang et al., proposed a multi-branch feature fusion convolutional neural network (MFF-CNN) to solve the difficult problems of multi-surface distribution of features and light sensitivity when automatically classifying automobile engine main bearing cover parts [24]. In addition, researchers have proposed improved Inception V3 [25] and Xception [26] for the identification of threaded connector parts.

The above algorithm can overcome the impact of illumination changes on recognition results to a certain extent; however, its research objective is relatively simple and cannot be effectively applied to the recognition of high-frequency workpieces that are complex within classes, have small gaps between classes, and have changeable postures. In order to effectively solve the problem of low recognition rate of high-frequency workpieces with the characteristics of complex intra-class and small gaps between classes under complex illumination changes, we propose in this paper a novel high-frequency workpiece image recognition model that uses EfficientNet-B1 [27] as the basic network and integrates multi-level network structures. The main contributions in this paper are threefold.

- (1) We introduce a lightweight mixed attention module (LMAM) to extract global workpiece image features with strong illumination robustness, and the global recognition results are obtained through the backbone network.
- (2) We use a weakly supervised area detection module to locate the locally important areas of the workpiece, which is then introduced into the branch network to obtain local recognition results.
- (3) We combine the global and local recognition results in the branch fusion module to achieve the final recognition of high-frequency workpiece images.

Experimental results on a high-frequency workpiece dataset made in the laboratory show that compared with various image recognition algorithms, the proposed algorithm has stronger adaptability to illumination changes and significantly improves the accuracy of high-frequency workpiece recognition.

The remainder of this paper is organized as follows: Section 2 outlines our method. Section 3 presents the experimental result of a high-frequency workpiece dataset made in the laboratory. Section 4 concludes this paper.

2. The Proposed Model

2.1. Overall Framework

The framework of the high-frequency workpiece recognition model proposed in this paper that integrates global attention and local attention is shown in Figure 1. It mainly consists of three modules: lightweight mixed attention module (LMAM), weakly supervised region detection module, and branch fusion module. First, the whole workpiece image I_1 is input into the lightweight mixed attention module (LMAM) to obtain enhanced features, and then, the multi-layer feature map M_g and recognition results P_1 are generated through the lightweight network EfficientNet-B1. Second, in the weakly supervised area detection module, the local workpiece image I_2 is intercepted based on the whole workpiece image I_1 and multi-layer feature map M_g , and it is imported into the LMAM and EfficientNet-B1 of another branch to obtain the recognition result P_2 . Finally, the branch fusion module is used to fuse the results of the two branches to obtain the final recognition result P .

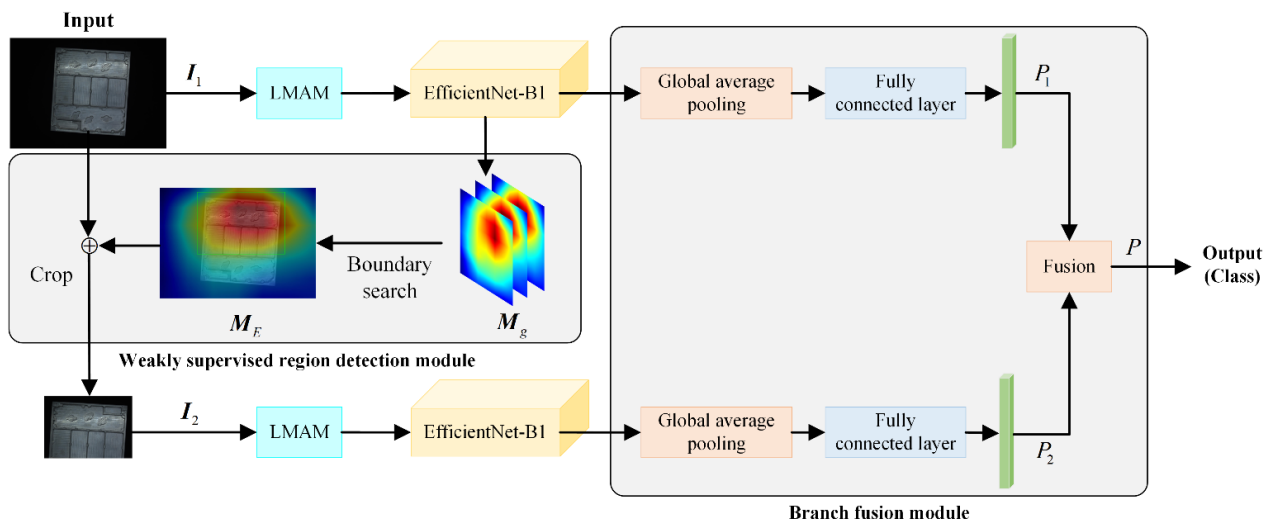


Figure 1. The framework of the proposed model.

2.2. The Lightweight Mixed Attention Module (LMAM)

In actual industrial production environments, the surface of a high-frequency workpiece is easily affected by factors such as uneven illumination and a large change in illumination, resulting in spots, shadows, insufficient light, and other phenomena in the collected workpiece images. If low-quality workpiece images are directly input into the network to extract features, effective image features cannot be obtained, and it is difficult to accurately identify the categories of high-frequency workpieces. In order to overcome the impact of the above interference information on workpiece recognition, this paper uses the LMAM module to enhance the features of the workpiece image, as shown in Figure 1.

Inspired by the convolutional block attention module (CBAM) [28] and efficient channel attention (ECA) [29], this paper proposes the LMAM shown in Figure 2 for preliminary screening. It uses the lightweight channel attention module (LCAM) and the lightweight spatial attention module (LSAM) to replace the channel attention module (CAM) and spatial attention module (SAM) in CBAM, respectively.

2.2.1. The Lightweight Channel Attention Module (LCAM)

Since the maximum pooling in CAM loses too much information, the LCAM in this article uses global standard deviation pooling to replace the global maximum pooling.

Then, one-dimensional convolution allows the network to focus on the learning of effective channels with less calculation, and its structure is shown in Figure 3.

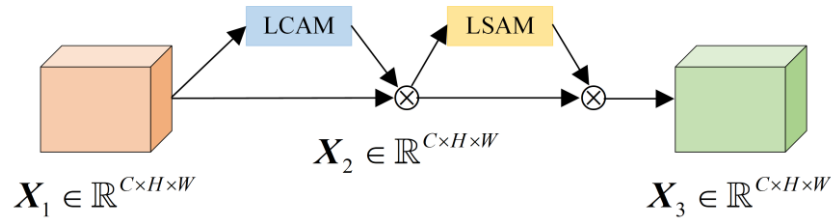


Figure 2. Structure diagram of LMAM.

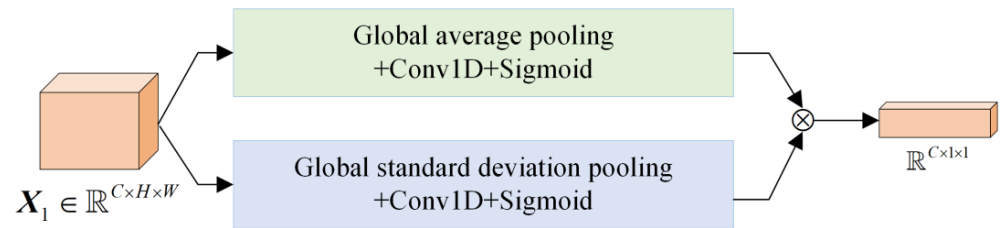


Figure 3. Structure diagram of LCAM.

Using global average pooling and global standard deviation pooling on the input feature map $X_1 \in \mathbb{R}^{C \times H \times W}$, the channel information description maps are obtained using the following two formulas, respectively.

$$z^m = \frac{1}{H \times W} \sum_{h=1}^H \sum_{w=1}^W X_1(h, w) \quad (1)$$

$$z^{sd} = \sqrt{\frac{1}{H \times W} \sum_{h=1}^H \sum_{w=1}^W (X_1(h, w) - z^m)^2} \quad (2)$$

where h and w are the coordinates in the height and width directions, respectively. H and W are the height and width of the image, respectively.

z^m and z^{sd} accumulate global information in different ways and then perform one-dimensional convolution on them, respectively. Note that z^m can effectively extract salient information, and z^{sd} can extract differential information. After a combination of convolution and activation, the channel attention weight is obtained.

$$M_c = \sigma(F_m^k(z^m)) \otimes \sigma(F_{sd}^k(z^{sd})) \quad (3)$$

where $F_m^k(\cdot)$ and $F_{sd}^k(\cdot)$ are one-dimensional convolution, respectively. k is the convolution kernel size. $\sigma(\cdot)$ is sigmoid function.

The convolution kernel size k can be adaptively calculated by the formula in the literature [18].

$$k = \psi(C) = \left\lceil \frac{\log_2(C) + b}{\gamma} \right\rceil_{odd} \quad (4)$$

where C is the number of channels, γ and b are set to 2 and 1, respectively. *odd* represents the nearest odd number.

2.2.2. The Lightweight Spatial Attention Module (LSAM)

The LSAM proposed in this paper is shown in Figure 4. In view of the small target difference between classes in the high-frequency workpiece dataset, a convolution kernel of size 3×3 is used to replace the convolution kernel of size 7×7 in SAM. At the same time, in order to obtain multi-scale context information, a dilated convolution with a kernel size

of 3×3 is added in parallel. Since dilated convolution will produce a grid effect, different receptive fields are summed and combined. There are only 18 parameters in the entire convolution process, which greatly reduces the number of parameters compared with the 49 parameters of SAM.

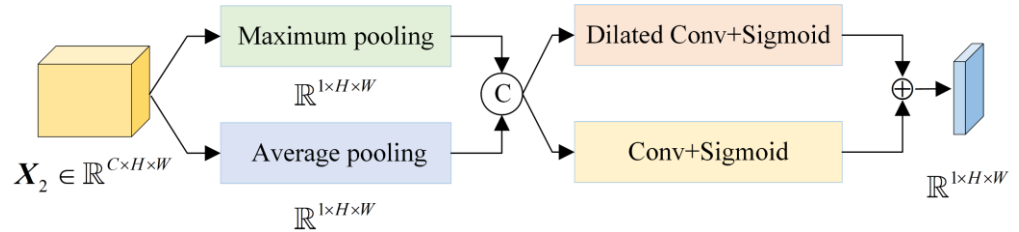


Figure 4. Structure diagram of LSAM.

Perform maximum pooling and average pooling on each pixel of the input feature map $X_2 \in \mathbb{R}^{C \times H \times W}$ by channel, then perform channel concatenation on the two obtained feature maps. Next, perform convolution with a kernel size of 3×3 and a dilated convolution with a kernel size of 3×3 , where the dilation rate is set to 1 and 2, respectively. In order to ensure that the size of the feature map remains unchanged, the padding parameters need to be set, and, finally, the spatial attention weight is obtained through activation and addition operations.

$$M_s(X_2) = \sigma\left(F_C^{3 \times 3}([P_{\max}(X_2); P_{\text{avg}}(X_2)])\right) \oplus \sigma\left(F_D^{3 \times 3}([P_{\max}(X_2); P_{\text{avg}}(X_2)])\right) \quad (5)$$

where $F_C^{3 \times 3}(\cdot)$ and $F_D^{3 \times 3}(\cdot)$ denote the convolution and the dilated convolution is a kernel size of 3×3 , respectively. $P_{\max}(\cdot)$ and $P_{\text{avg}}(\cdot)$ denote the maximum pooling and average pooling, respectively.

2.3. The Weakly Supervised Region Detection Module

High-frequency workpieces have the characteristics of many types and small differences between classes, and the small differences between workpieces often appear in specific local areas. Based on these characteristics, this paper proposes to use a weakly supervised region detection module, which includes two mechanisms, boundary search and cropping, to locate areas with significant differences in the workpiece image, as shown in Figure 1.

The global multi-layer feature map M_g of the input image is generated through the backbone network EfficientNet-B1, and the energy map M_E is obtained by superimposing the feature maps of all channels. In order to eliminate the interference of negative elements, all elements of M_E are normalized to $[0, 1]$ to obtain the scaled energy map.

$$\hat{M}_E(i) = \frac{M_E(i) - \min(M_E)}{\max(M_E) - \min(M_E)} \quad (6)$$

where $\max(M_E)$ and $\min(M_E)$ denote the maximum value and the minimum value in $M_E(i)$, respectively.

In order to further improve the positioning accuracy, bilinear interpolation is used to upsample $\hat{M}_E$ to the size of 25×25 .

$$\bar{M}_E = \text{bilinear}(\hat{M}_E) \quad (7)$$

Then, $\bar{M}_E$ is aggregated into two one-dimensional structured energy vectors.

$$\begin{cases} V_w = \sum_{i=0}^H \overline{M}_E(i, W) \\ V_h = \sum_{j=0}^W \overline{M}_E(H, j) \end{cases} \quad (8)$$

where V_w and V_h denote one-dimensional structured energy vectors along the width and height directions of space, respectively.

Taking V_w as an example, extract the energy of different elements.

$$\begin{cases} E[0 : W] = \sum_{i=0}^W V(i) \\ E[x_1 : x_2] = \sum_{i=x_1}^{x_2} V(i) \end{cases} \quad (9)$$

where $E[0 : W]$ denotes the energy sum of all elements in the width vector, and $E[x_1 : x_2]$ denotes the energy of the area with width from x_1 to x_2 .

Define the key area in the global image as occupying the smallest area and meeting the following condition:

$$\begin{cases} E[x_1 : x_2] / E[0 : W] > \gamma \\ E[y_1 : y_2] / E[0 : H] > \gamma \end{cases} \quad (10)$$

where γ denotes the preset threshold.

The width boundary coordinate $[x_1 : x_2]$ and the height boundary coordinate $[y_1 : y_2]$ of the area can be found automatically by using the boundary search mechanism. Then, a cropping mechanism is used to intercept effective workpiece information and valuable background information from the original image according to the entire boundary coordinate $[x_1 : x_2, y_1 : y_2]$, and the local workpiece image I_2 is finally obtained.

2.4. The Branch Fusion Module

In order to simultaneously utilize the global information and local information of the image and weigh the role of the two types of information in workpiece image recognition, a branch fusion module is proposed to take into account the recognition results of the two branches to further improve the accuracy of the recognition, as shown in Figure 1. It should be noted that in the model proposed in this paper, the two channel attention modules do not share parameters to extract workpiece features at different scales. The calculation formula of the fused recognition score is as follows:

$$P = \mu P_1 + (1 - \mu) P_2 \quad (11)$$

where P_1 and P_2 denote the recognition result of the global image and the recognition result of the local image, respectively. And μ denotes the balancing factor that weighs the recognition results of different branches.

3. Experimental Results

3.1. Dataset

The high-frequency workpiece image dataset used in the experiment comes from a research institute in China. The dataset contains 20 different categories of workpieces. Each category of workpieces has 1000 images, totaling 20,000 images. The size of the image is 3822×2702 . The dataset is randomly divided into a training set and a validation set in a ratio of 7:3. Figure 5 shows examples of six different categories of high-frequency workpieces, where the red circle indicates the small differences between the six categories. It can be seen from Figure 5 that for each type of high-frequency workpiece image, its internal texture is characterized by complex texture; for different types of high-frequency workpiece images, the differences between the classes are small. In addition, there are

also local bright spots caused by uneven lighting in the image. Therefore, classifying multi-category high-frequency workpiece images is an equally challenging task.



Figure 5. Some high-frequency workpiece image samples.

3.2. Experimental Settings

The training process of ML-Efficient-B1 is carried out by the Adam optimizer [30], where the initial learning rate is set to 1×10^{-4} . The batch size and image size are set to 8 and 224×224 , respectively. The training process is performed until 50 epochs using a machine with NVIDIA GeForce GTX 1660 SUPER GPU (NVIDIA is located in Santa Clara, California, USA). The implementation is carried out using the PyTorch1.8 package [31]. Note that the weight parameter initialization of the network is obtained by using pre-training on ImageNet.

3.3. Model Parameter Selection

The cropping range threshold γ determines the size of the effective area extracted from the global image, further affecting the accuracy of workpiece recognition. If γ is too small, excessive loss of workpiece features will result. On the contrary, if γ is too large, the network will not be able to focus on important local features. Therefore, the intercepted area should be limited to a reasonable range. This article limits it to $[0.60, 0.80]$. In addition, the parameter μ in Equation (11) also greatly affects the network's emphasis on different branches.

In order to determine the optimal threshold, the balance factor was first set to 0.6, and then the recognition accuracy of high-frequency workpieces was tested when the values of γ were 0.6, 0.65, 0.70, 0.75, and 0.80, respectively. The experimental results are shown in Table 1.

Table 1. The impact of the size of threshold γ on the recognition results.

The Values of γ	Accuracy (%)
0.60	96.5
0.65	97.7
0.70	98.3
0.75	97.5
0.80	96.8

As can be seen from Table 1, as the cropping range threshold increases, the accuracy first increases and then decreases. When the threshold is 0.70, the algorithm in this paper achieves the best performance. Therefore, 0.70 is selected as the final threshold.

In order to evaluate the impact of the balance factor μ on the recognition results, first γ is fixed at 0.70, and then the high-frequency workpiece recognition accuracy is tested when μ takes different values. The experimental results are shown in Table 2.

Table 2. The impact of balancing factors on recognition results.

The Values of μ	Accuracy (%)
0.0	96.7
0.1	97.2
0.2	97.4
0.3	97.6
0.4	97.7
0.5	97.8
0.6	98.3
0.7	97.5
0.8	97.0
0.9	96.3
1.0	96.1

As can be seen from Table 2, when μ is 0.6, the high-frequency workpiece image obtains the highest recognition accuracy. As μ decreases, the recognition accuracy of high-frequency artifacts gradually decreases. This is because the cropped partial image contains less information, and relying too much on this part will cause the network to be unable to obtain better results. Therefore, we take $\mu = 0.6$ as the balance parameter.

3.4. Comparison of Recognition Performance

In order to verify the effectiveness of the high-frequency workpiece image recognition model proposed in this paper, a comparative experiment was conducted with a variety of image recognition algorithms. Comparison algorithms include RTMM [20], JLS-DL [23], EfficientNet [27], a mechanical parts identification algorithm based on convolutional neural network [32] (denoted as WorkNet-2), main bearing cover parts identification based on deep learning [24] (denoted as MFF-CNN), the part recognition algorithm based on improved convolutional neural network [33] (denoted as Xception-P), NOAH [34], and RAFIC [35]. The experimental results are shown in Table 3.

Table 3. Recognition results of different algorithms.

Method	The Total Number of Images	The Number of Correct Recognitions	Accuracy (%)
EfficientNet [27]	6000	5142	86.2
WorkNet-2 [32]	6000	5370	90.0
MFF-CNN [24]	6000	5424	90.9
Xception-P [33]	6000	5496	92.1
RTMM [20]	6000	5586	93.6
JLS-DL [23]	6000	5664	94.9
NOAH [34]	6000	5763	96.1
RAVIC [35]	6000	5829	97.3
ML-EfficientNet-B1 (Our model)	6000	5868	98.3

It is shown in Table 3 that the recognition performance of the proposed ML-EfficientNet-B1 for high-frequency workpiece images is significantly better than other comparison methods. Specifically, the increases over that of the EfficientNet, WorkNet-2, MFF-CNN, Xception-P, RTMM, JLD-DL, NOAH, and RAFIC methods on the high-frequency workpiece image dataset are 12.1%, 8.3%, 7.4%, 6.2%, 4.7%, 3.4%, 2.2%, and 1.0%, respectively.

In addition, Figure 6 shows the confusion matrix of the recognition results of the proposed model. As can be seen from Figure 6, the recognition accuracy of the ML-EfficientNet-B1 for each type of high-frequency workpiece is above 90%.

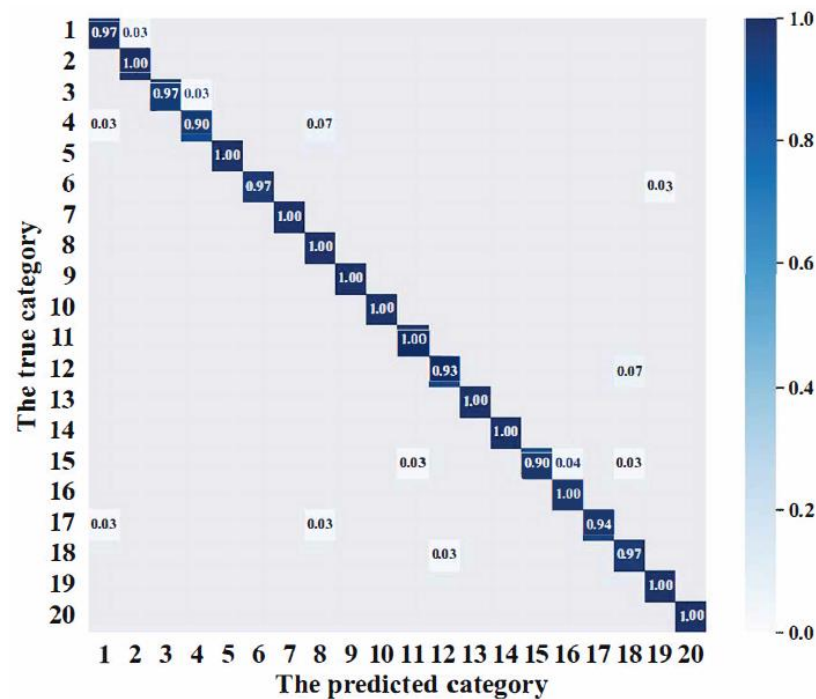


Figure 6. Confusion matrix on the high-frequency workpiece image dataset.

3.5. Ablation Study

In this sub-section, we carry out an ablation study of the proposed ML-EfficientNet-B1 in order to show the new modules on improving the network performance. Specifically, based on retaining the EfficientNet-B1 network, the effect of the lightweight mixed attention module (LMAM), the weakly supervised region detection module (WSRDM), and the branch fusion module (BFM) on the entire model is verified by controlling variables. The experimental results are shown in Table 4.

Table 4. The impact of different modules on network performance.

LMAM	WSRDM	BFM	Accuracy (%)
×	×	×	86.2
✓	×	×	96.7
✓	✓	×	97.4
✓	✓	✓	98.3

× means that the module is not used; ✓ means that the module is used.

The following can be seen from Table 4: (1) When directly using the EfficientNet-B1 network to classify high-frequency workpiece images, the recognition accuracy is 86.2%. (2) After adding the LMAM module based on the backbone network EfficientNet-B1, the accuracy rate is increased by 10.5%. This shows that the LMAM module has the ability to perceive feature information of different color channels, overcome the impact of illumination changes, and more effectively extract the features of high-frequency workpiece images. (3) On the basis of using the LMAM module, with further added WSRDM, the accuracy rate further increased by 0.7%, which shows that the feature learning of the WSRDM focuses on differentiated effective areas, which can improve the accuracy of recognition. (4) When LMAM, WSRDM, and WSRDM are used simultaneously, the accuracy rate is further improved by 0.9%, indicating that combining global information and local information can improve the recognition performance of high-frequency workpieces.

3.6. Visualization Results

In order to verify the effectiveness of the proposed algorithm, some images are randomly selected for visual display on the same test set. Figure 7 shows the subtle differences between different categories of high-frequency workpieces and the visualization of attention features before and after algorithm improvement.

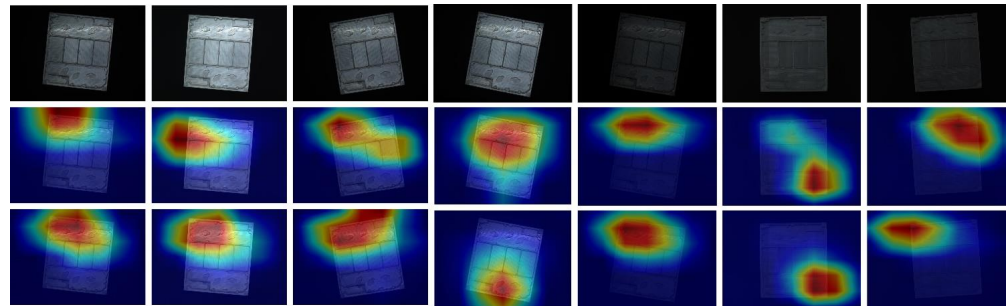


Figure 7. Attention feature visualization. Note that the first row denotes the original images, the second row denotes the results obtained by EfficientNet, and the third row denotes the results obtained by the proposed ML-EfficientNet-B1.

As can be seen from Figure 7, compared with the original EfficientNet network, the improved ML-EfficientNet-B1 model makes the network more focused on the area where the boss is located. Therefore, the proposed EfficientNet-B1 model can extract more discriminative workpiece image features, significantly improving the recognition performance of the network.

Table 5 demonstrates the inference time of the proposed ML-EfficientNet-B1 and the other algorithms on a machine with NVIDIA GeForce RTX 3090 GPU, for an image with the spatial resolution of 224×224 . As seen from this table, the inference time of the proposed network is 0.133 (s), which is acceptable for providing a high performance.

Table 5. Complexity of different algorithms.

Method	Execution Time
EfficientNet [27]	0.082 (s)
WorkNet-2 [32]	0.051 (s)
MFF-CNN [24]	0.104 (s)
Xception-P [33]	1.320 (s)
RTMM [20]	30.517 (s)
JLS-DL [23]	0.208 (s)
NOAH [34]	0.164 (s)
RAFIC [35]	0.185 (s)
ML-EfficientNet-B1 (Our model)	0.135 (s)

3.7. Discussion and Limitation

In summary, the improved model proposed in this paper can effectively improve the recognition accuracy of multi-category high-frequency workpiece images, effectively solve the impact of illumination changes on high-frequency workpiece recognition results, and make full use of the global and local features of workpiece images. When the model in this paper is used to classify high-frequency workpiece images, it only uses the top surface image and does not consider the thickness information of the workpiece. It cannot be applied to the classification of high-frequency workpiece images with different thicknesses but the same top surface information. In subsequent algorithms, combining multi-view images can be considered to improve classification accuracy and the number of categories.

4. Conclusions

This paper proposes a high-frequency workpiece image recognition model that integrates multi-level network structures. First, LMAM is introduced to enhance the feature extraction capability of the network and reduce the impact of illumination changes on high-frequency artifact recognition results. Then, a weakly supervised area detection module is used to search and locate differentiated local images. Finally, the branch fusion module is used to effectively balance the network's ability to capture global features and local features of the workpiece images. Experimental results show that compared with the original EfficientNet network, the model in this paper improves the recognition accuracy of high-frequency artifacts by 12.1%. In addition, compared with various image recognition methods, the model in this paper has a certain degree of improvement in the recognition accuracy of high-frequency workpieces. In the future, in order to better arrange it on the production line, we will study a more lightweight high-frequency workpiece classification model.

Author Contributions: Conceptualization, Y.O. and C.S.; methodology, Y.O.; software, Y.O. and C.S.; validation, Y.O.; writing—original draft preparation, Y.O.; writing—review and editing, J.L. and R.Y.; visualization, C.S.; supervision, J.L.; funding acquisition, R.Y. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported by the National Natural Foundation of China, grant No. 52175130.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The dataset that was generated and analyzed during this study is available from the corresponding author upon reasonable request, but restrictions apply to data reproducibility and commercially confident details.

Acknowledgments: The authors gratefully acknowledge the useful comments of the reviewers.

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

References

1. Wübbeke, J.; Meissner, M.; Zenglein, M.J.; Ives, J.; Conrad, B. Made in China 2025. *Mercat. Inst. China Studies. Pap. China* **2016**, *2*, 4.
2. Zenglein, M.J.; Holzmann, A. Evolving made in China 2025. *MERICS Pap. China* **2018**, *8*, 78.
3. Zhou, J. Intelligent Manufacturing—Main Direction of “Made in China 2025”. *China Mech. Eng.* **2015**, *26*, 2273.
4. Zhao, H.; Jia, J.; Koltun, V. Exploring self-attention for image recognition. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Seattle, WA, USA, 13–19 June 2020; pp. 10076–10085.
5. Sampurno, R.M.; Liu, Z.; Abeyrathna, R.M.; Ahamed, T. Intrarow Uncut Weed Detection Using You-Only-Look-Once Instance Segmentation for Orchard Plantations. *Sensors* **2024**, *24*, 893. [\[CrossRef\]](#) [\[PubMed\]](#)
6. Alam, L.; Kehtarnavaz, N. Improving Recognition of Defective Epoxy Images in Integrated Circuit Manufacturing by Data Augmentation. *Sensors* **2024**, *24*, 738. [\[CrossRef\]](#) [\[PubMed\]](#)
7. Sheykhmousa, M.; Mahdianpari, M.; Ghanbari, H.; Ghamisi, P.; Homayouni, S. Support vector machine versus random forest for remote sensing image classification: A meta-analysis and systematic review. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2020**, *13*, 6308–6325. [\[CrossRef\]](#)
8. Zhao, C.; Qin, Y.; Zhang, B. Adversarially Learning Occlusions by Backpropagation for Face Recognition. *Sensors* **2023**, *23*, 8559. [\[CrossRef\]](#)
9. Wang, H.; Ma, L. Image Generation and Recognition Technology Based on Attention Residual GAN. *IEEE Access* **2023**, *11*, 61855–61865. [\[CrossRef\]](#)
10. Yi, X.; Qian, C.; Wu, P.; Maponde, B.T.; Jiang, T.; Ge, W. Research on Fine-Grained Image Recognition of Birds Based on Improved YOLOv5. *Sensors* **2023**, *23*, 8204. [\[CrossRef\]](#)
11. Li, Z.; Tang, H.; Peng, Z.; Qi, G.; Tang, J. Knowledge-guided semantic transfer network for few-shot image recognition. *IEEE Trans. Neural Netw. Learn. Syst.* **2023**. [\[CrossRef\]](#)
12. Xu, C.; Gao, W.; Li, T.; Bai, N.; Li, G.; Zhang, Y. Teacher-student collaborative knowledge distillation for image classification. *Appl. Intell.* **2023**, *53*, 1997–2009. [\[CrossRef\]](#)
13. Yao, P.; Wu, H.; Gao, B.; Tang, J.; Zhang, Q.; Zhang, W.; Yang, J.J.; Qian, H. Fully hardware-implemented memristor convolutional neural network. *Nature* **2020**, *577*, 641–646. [\[CrossRef\]](#) [\[PubMed\]](#)

14. Zhou, G.; Li, J.; Song, Q.; Wang, L.; Ren, Z.; Sun, B.; Hu, X.; Wang, W.; Xu, G.; Chen, X.; et al. Full hardware implementation of neuromorphic visual system based on multimodal optoelectronic resistive memory arrays for versatile image processing. *Nat. Commun.* **2023**, *14*, 8489. [\[CrossRef\]](#) [\[PubMed\]](#)
15. Ou, Y.; Luo, J.; Li, B.; He, B. A classification model of railway fasteners based on computer vision. *Neural Comput. Appl.* **2019**, *31*, 9307–9319. [\[CrossRef\]](#)
16. Luo, J.; He, B.; Ou, Y.; Li, B.; Wang, K. Topic-based label distribution learning to exploit label ambiguity for scene classification. *Neural Comput. Appl.* **2021**, *33*, 16181–16196. [\[CrossRef\]](#)
17. Luo, J.; Wang, Y.; Ou, Y.; He, B.; Li, B. Neighbour-based label distribution learning to model label ambiguity for aerial scene classification. *Remote Sens.* **2021**, *13*, 755. [\[CrossRef\]](#)
18. Xu, Y.; Feng, K.; Yan, X.; Yan, R.; Ni, Q.; Sun, B.; Lei, Z.; Zhang, Y.; Liu, Z. CFCNN: A novel convolutional fusion framework for collaborative fault identification of rotating machinery. *Inf. Fusion* **2023**, *95*, 1–16. [\[CrossRef\]](#)
19. Xu, Y.; Yan, X.; Sun, B.; Feng, K.; Kou, L.; Chen, Y.; Li, Y.; Chen, H.; Tian, E.; Ni, Q.; et al. Online Knowledge Distillation Based Multiscale Threshold Denoising Networks for Fault Diagnosis of Transmission Systems. *IEEE Trans. Transp. Electr.* **2023**. [\[CrossRef\]](#)
20. Xu, W.; Li, B.; Ou, Y.; Luo, J. Recognition algorithm for metal parts based on ring template matching. *Transducer Microsyst. Technol.* **2021**, *40*, 128–131.
21. Yin, K.; Ou, Y.; Li, B.; Lin, D. Fast identification algorithm of high frequency components based on ring segmentation. *Mach. Des. Manuf.* **2022**, *12*, 196–200.
22. Wang, Y.; Chen, H.; Zhao, K.; Zhao, P. A mechanical part sorting method based on fast template matching. In Proceedings of the 2018 IEEE International Conference on Mechatronics, Robotics and Automation (ICMRA), Changchun, China, 5–8 August 2018; pp. 135–140.
23. Yang, T.; Ou, Y.; Su, X.; Wu, X.; Li, B. High frequency workpiece deep learning recognition algorithm based on joint loss supervision. *Mach. Build. Autom.* **2023**, *52*, 30–33.
24. Zhang, P.; Shi, Z.; Li, X.; Ouyang, X. Classification algorithm of main bearing cap based on deep learning. *J. Graph.* **2021**, *42*, 572–580.
25. Szegedy, C.; Vanhoucke, V.; Loffe, S.; Shlens, J.; Wojna, Z. Rethinking the inception architecture for computer vision. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA, 26 June–1 July 2016; pp. 2818–2826.
26. Chollet, F. Xception: Deep learning with depthwise separable convolutions. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, HI, USA, 21–26 July 2017; pp. 1800–1807.
27. Tan, M.; Le, Q. Efficientnet: Rethinking model scaling for convolutional neural networks. In Proceedings of the International Conference on Machine Learning (ICML), Long Beach, CA, USA, 9–15 June 2019; pp. 6105–6114.
28. Woo, S.; Park, J.; Lee, J.Y.; Kweon, I. Cbam: Convolutional block attention module. In Proceedings of the European Conference on Computer Vision (ECCV), Munich, Germany, 8–14 September 2018; pp. 3–19.
29. 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/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Seattle, WA, USA, 13–19 June 2020; pp. 11534–11542.
30. Kingma, D.P.; Ba, J.L. Adam: A method for stochastic optimization. *arXiv* **2014**, arXiv:1412.6980.
31. Paszke, A.; Gross, S.; Massa, F.; Lerer, A.; Bradbury, J.; Chanan, G.; Killeen, T.; Lin, Z.; Gimelshein, N.; Antiga, L.; et al. Pytorch: An imperative style, high-performance deep learning library. *Adv. Neural Inf. Process. Syst.* **2019**, *32*. [\[CrossRef\]](#)
32. Duan, S.; Yin, C.; Liu, M. Recognition Algorithm Based on Convolution Neural Network for the Mechanical Parts. *Adv. Manuf. Autom.* **2019**, *484*, 337–347.
33. Yang, L.; Gan, Z.; Li, Y. Part recognition based on improved convolution neural network. *Instrum. Tech. Sens.* **2022**, *5*, 82–87.
34. Li, C.; Zhou, A.; Yao, A. NOAH: Learning Pairwise Object Category Attentions for Image Classification. *arXiv* **2024**, arXiv:2402.02377.
35. Lin, H.; Miao, L.; Ziai, A. RAFIC: Retrieval-Augmented Few-shot Image Classification. *arXiv* **2023**, arXiv:2312.06868.

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.