Research Article

Utilization of Both Machine Vision and Robotics Technologies in Assisting Quality Inspection and Testing

Jiagui Tao, Qingmiao Chen, Jinxia Xu, Heng Zhao, and Siqi Song

State Grid Jiangsu Electric Power Co., Ltd. Research Institute, Nanjing 211103, China
Wuhan NARI Co. Ltd., State Grid Electric Power Research Institute, Wuhan 430074, China

Correspondence should be addressed to Qingmiao Chen; cqmwuhannari@163.com

Received 7 January 2022; Revised 9 February 2022; Accepted 12 February 2022; Published 16 March 2022

Abstract

Inspecting the smoothness and durability of the exterior and interior surfaces of manufactured products such as large passenger aircraft requires utilizing manufacturing precision instruments. Manual techniques are mainly adopted in inspections, which are, however, much costly and have low efficiency. Moreover, manual operation is more prone to missing and false inspections. To cope with these issues, effective precision instruments have been needed to devise intelligent inspections for impalpable damages as necessary tools. When many inspection tasks are investigated, very few public datasets are available to recognize the intelligent inspection of precision instruments. On the other hand, precision instruments can be applied to a wide variety of damages. YOLO V3 was proposed based on acquiring and processing the image appearances, which deal with the surface inspection of the precision instrument in this study. The image dataset of the impalpable damage was initially established. More specifically, the YOLO V3 detection network is leveraged to roughly calculate the location of the damage appearances and identify the damage type. Afterward, the designed level set algorithm was employed to obtain more accurate damage locations in the image block utilizing the characteristics of different types of damages. Finally, a quantitative analysis was performed employing the refined detection results. A deep architecture that can intelligently conduct damage detection was proposed. Besides, the proposed method exhibited a strong tolerance for unknown types of damages and excellent flexibility and adaptability. Extensive experimental results demonstrated that the proposed method can alleviate the shortcomings of the traditional inspection methods. Thus, it provided technical guidance for the applications of the intelligent orbital inspection of the robots. The proposed method with a more precise and noninvasive inspection technique remarkably accelerates the inspection time of impalpable damages on surfaces.

1. Introduction

The manufacturing industry is the core of any national economy. Although China is a major manufacturer across the world, the qualities of various manufactured products and semimanufactured products are far behind the ones that are manufactured by developed countries such as the United States. In particular, there is a lack of surface quality, namely, the smoothness and precision of various important components in steel manufacturers. The surface quality of the main steel products such as automobile plates, home appliance plates, and decorative stainless steel plates plays an indispensable role in the assessment of the overall quality of manufactured products and can even affect their competitiveness in the market. By leveraging the traditional manual and visual inspection techniques to examine surface qualities of steel plates, it is highly probable to generate false-positive and wrong-negative detections related to surface defects such as the so-called “visual fatigue.” This is the case in many steel companies that try to cope with when the quality issues of the steel surface are a concern. This problem directly hurts economic benefits. Instead of conducting manual inspection techniques, for example, for the parts of aircraft by manufacturers, machine vision techniques can be utilized to detect those. Noticeably, the method of intelligently detecting and localizing the impalpable damages of the aircraft has the advantages of high efficiency, low cost, and
does not include the interventions of human actions. However, more advanced techniques could be possibly needed to detect impalpable damages on the surfaces of the manufactured products, for example, machine vision and robotics applications concerning deep architecture tools.

A systematic description of a deep architecture can be briefed as follows. The hardware of the proposed method is mainly composed of a complementary metal oxide semiconductor camera (CMOS), lens, light source, and computer. More specifically, the CMOS cameras and lenses are responsible for real-time image acquisition that can convert optical images into digital signals featuring of small size, lightweight, and hardly influenced by external interference. Therefore, it is pervasively used in industrial visual inspections. When concurrently considering the detection accuracy and the working distance, the Gray Dot Company’s industrial camera called the Blackfly PoE GigE Color Camera (BFLY-PGE-31S4C-C) model is utilized. We choose the appropriate lens regarding the selected camera model and the size of the field view. The interface of the selected camera is based on C language; thus, the resolution size is fixed to 2048 × 1536. The surface size of the sensor target is assigned to 1/8″ in the implementation, and the size of the aluminum foil in the actual test is set to approximately 9 mm. Thus, the maximum target surface is set to 1/8″ and a fixed-focus lens with a focal length of 25 mm is utilized. The light source is an important module in the implementations and extensively determines the qualities of captured images [1, 2]. A selection of a successful lighting source can substantially reduce the difficulty of the subsequent image processing. When considering many factors such as response speed, life span, and stability, the light source called light-emitting diode (LED) is selected in the proposed method. When the front side is used to highlight the wrinkles on the surface of the aluminum foil, a ring light source with uniform light is utilized. Besides, we incorporate a light source on the backside to make the edge of the outer contour of the aluminum foil more prominent.

The visual inspection software primarily conducts image processing on the pictures acquired by the camera based on semantically analyzed results in the proposed method [3], and the interface of the visual inspection software is implemented by QT software under the VS2015 platform. To crunch the outputs of the processed pictures from the camera in real time, the software adopts a multithread mode. The main thread is responsible for the interface display of the entire software and controls the opening and closing of the second thread in the used software. The main interface includes the menu bar, image displaying area, and a display area. The menu bar contains the buttons with six different functions presented in Table 1.

The function of the second thread is to conduct the calculation of the camera communication and image processing. The camera communication module is mainly responsible for the image transmissions between the camera and the software. When the “turn on the camera” function button is activated in the main interface, the second thread gets started and the picture is read by the camera. Subsequently, the picture is transmitted to the image processing module that carries out different types of image processing tasks such as filtering, threshold segmentation, and edge extraction and, then, subsequently displays the extracted results in the image displaying area of the main thread. Thus, the obtained parameters are transmitted to the calculation of the camera module. Hence, the calculation module finalized the process.

| Function ID | Details                          |
|-------------|----------------------------------|
| A           | Turn on camera                   |
| B           | Camera calibration               |
| C           | Save calibration result          |
| D           | Read calibration file            |
| E           | Save current picture             |
| F           | Read local picture               |

In this study, we investigate the visual processing algorithm based on YOLO V3, which primarily resolves the problems of deep learning-based damage detections regarding the types of known damages to detect and analyze unknown damages by utilizing the extracted knowledge. Empirical evaluations have shown that the key techniques based on the conventionally conducted manual and visual inspections have some sort of weakness. On the other hand, the proposed method can remarkably accelerate the inspection time of impalpable damages of surfaces. Moreover, the proposed method can substantially improve the maintenance and inspection efficiency of the manufactured products by utilizing a few parameters.

With the combination of the charge-coupled device (CCD)-based applications and machine vision techniques, we propose a more precise and noninvasive inspection technique that can be leveraged to detect the surface quality of various products. The proposed method can seamlessly combine the high-speed CCD cameras with computer vision systems that determine the necessity of discovering deep visual features in visual quality prediction. Then, it can timely provide both detected images and accurate feedback. Experiments have shown that the proposed method has become an effective tool to inspect the surface quality of manufactures products, for example, steel plates.

The rest of the manuscript is organized as follows: Section 2 presents the related work mentioning the necessary hardware, software, and configurations to extract images and its digitalization process that leads to conducting machine vision. Section 3 presents one of the deep learning methods called the convolutional neural network and its fundamental structure. Section 4 introduces the proposed model with an experimental setup. Section 5 concludes the research.

2. Related Work

2.1. Machine Vision Hardware. Machine vision is a comprehensive concept including many techniques such as image processing, mechanical engineering technology, electric lighting source, optical imaging, digital video technology, computer software, and hardware technology [3]. An application of a typical machine vision involves
image capturing, setting up a system of a light source, image digitization, digital image processing, intelligent decision-making, and execution of mechanical control. The current machine vision platform typically includes the following components: NVIDIA VisionWorks, Recognition, and Open Visual Inferencing and Neural Network Optimization (Open VINO). When the hardware cost and the interface difficulty of the platform operation were a concern, many systems leveraged the Open VINO as the development kit of the computer vision to process the captured images. Hence, it significantly impacted on the system performance. The platform requirements were relatively low when compared to other computer vision toolkits. Herein, we chose the Windows 10 with Intel 10-th Corel CPU. The video-capturing device was chosen as a handheld internet protocol (IP) camera. Network cameras were a modern stock of products that agreed with traditional cameras and reticulation video technology [1, 3, 4]. The apprehension of the video foreshadowing by the camera was a finger and compacted by a high-effectiveness compression turning. The detention video was transmitted to the web salver through the reticulation bus bar. Users on the plexus could use the browser to instantly invigilate the camera shown on the web salver. Meanwhile, the accredited users could also superintend the actions called result/incline crystalline of the camera or the exercise of the system configuration. IP fret camera was a digital scheme supported by the plexus transmission. The reticulation camera also had a netting interface production regarding the common compound of extraordinary video production utilizing the British National Corpus interface (BNC). Hence, this could instantaneously associate the camera to the topical local area network (LAN) for video transmission.

2.2. Deep Learning for Visual Modeling. A convolutional neural network (CNN) was proposed by [5]. When the description of the shoal show is under investigation, the CNN appears to be an upper hand when in the case of inputting the meshwork as a conception or video. Each resemblance/video can be instantaneously placed as the input of the mesh in a dusky-spar journey in this text by voiding the obstacle of the old-fashioned optics modeling. The obscure erudition-supported shape essence and optical reconstruction have powerful advantages when manipulating idol/videos. The flexure can automatically catch pigment, interweave, conceive, and topological structure in the conception characteristics [6–9]. The mortally blue-stocking form can be customized for uncertain optic labor. AlexNet is one of the most far-manner CNN nets to process the information of the system apparition, which not only makes the better authority of the CNN in the optic modeling but also aids the expanse of sagacious scholarship in address notice, the outgrowth of the native style, and the reinforcement lore [3]. The AlexNet issues convolutional neural reticulum to end and rank picture characteristics. The mesh form has a septenary stratum, comprehending a five course of convolutional coping and two seam of plenteously joined seam [7]. It relatively inserts the rectified linear input (ReLU) and dropout for data augmentation and pooling. This cunning standard has five convolutional copings. Three convolutional couches are adjuncts to the greatest pooling course. From the structure of AlexNet, we can completely save AlexNet. The first footing is expressed as follows: the dimension of the volute nucleus is set to $11 \times 11$, the pace bigness is set to $4 \times 4$, and the scalar of the ravine is set to $96$, which is succeeded by a real activating sine and provincial regularization. Then, the highest pooling couch with a volume nucleus dimension of $3 \times 3$ and a measured adjustment of $2 \times 2$ is ultimately incorporated. The subordinate couch can be shown as syn. The adjustment of the volute nucleus is set to $5 \times 5$, the action largeness is set to $1 \times 1$, the amount of passage is set to 256, and the last is the highest pooling sill with six-volume nucleus of $3 \times 3$ and a proceeding adjustment of $2 \times 2$. The third footing, called quarten course, is set to 1/5 stratum: the gauge of the volute nucleus is set to $3 \times 3$, the footstep dimension is set to $1 \times 1$, and the enumeration of the ravine is partially 384, 384, and 256. After the convolutional coping, a pooling coping is typically incorporated. The pooling stratum can completely impair the gauge of the die, thereby, reducing the parameters in the extreme bond lift. Using the pooling lift can advance the summation of the characteristics and hinder the over-becoming proposition. The function of the fully connected layer optimally conducts the image classification. Different neural networks have different layers of the entire network, but the functions are the same in the practice. The number of neurons connected to the fully connected layer is greatly reduced after being processed by the convolutional layer and the pooling layer. Afterward, the final softmax output can be determined regarding the number of classification labels.

2.3. The Prediction Model of Visual Quality. A video server can be examined as a plot of ranking the inventive videos [10–14]. Thus, the assessment of the video characteristics mainly turns to the assessment of the independent fallacy, in which the character of each settle is an estimate and the musty sarcastic video can be held by amalgamating the qualities of these rules. Peak signal-to-noise ratio (PSNR) is widely utilized for the disposition assessment of the videos. However, it is ignorant to calculate the source of the video’s crookedness. However, the calculation of the PSNR is possessed on a prominent pixel dispute. Thus, the delusion remissness of reflection comments on the score of the visibility of the crookedness. Therefore, the PSNR cannot always muse the unbiased mortal taste. In the transmission system of the web real-time communication (WebRTC) back videos, the rank assessment of no-advertent video acquires a sizable party in video transmission. Reference [8] proposed to alloy the fame of the variegated and adjusted distortions of the measured video apart. Then, the calling ratio of the probative video is multiplied by the design distortions. Reference [10] suggested the four pure parameters that an idol video describes to break down rubbish video profit. Reference [10] introduced an unnatural direct framework of no-deference video over an obstruct network. The opposite look aims to standardize the shut of treatise
losses of this kind. Reference [11] examined a divisive framework and quality profit without specification (QoS) feed to psychoanalyze the dash of network on video avowal. Reference [12] fitted a noise reduction (NR) video that overlooks the NORM system toward H.264/AVC-coded video to anatomize the boom of swell errors on videos. In the macroblock equitable, NORM contributes a valuation of the maxillary skeletal expansion (MSE) crookedness. NORM had a generous linear reciprocation with the crookedness dress by the full reference (FR) methods.

Figure 1 shows the defect detection of the proposed framework that consists of three modules: (1) digital image acquisition, (2) embedded system for image preprocessing, and (3) defect detection based on deep learning.

3. The Proposed Method and Experiments

The proposed method consists of three modules, namely, image acquisition system, image processor, and convolutional neural network that defines the system.

**Module 1.** The image acquisition system is an important part of the online quality inspection systems for the surface of the manufactured product, which functions as the “eye” to the system. According to the machine vision-based surface quality inspection system, the proposed method intelligently selects a top-performance linear array charged-coupled device camera (CCD) or an area array CCD camera and presents the comparison between linear array CCD camera and an area array CCD camera [4]. Then, the system is applied to detect the surface quality of the plate on the cold rolling production line. The role of the linear array high-speed CCD camera is the sensor of the system. The Gigabit Ethernet camera (GigE) including a high-speed line scan camera leverages the CCD technique. Its speed is consistent with the Ethernet and the frame representation format. The GigE camera generates frame data, wherein each frame is divided into multiple Ethernet packets during transmission. The frame is composed of the data header “Leader,” data load “Payload,” and data tail “Trailer.” Each image is delivered to the data payload. The bandwidth of the camera is determined by the number of transmitted bytes and the transmission time. The number of bytes of each X bytes/frame is then calculated.

The payload is the image information; the package is the size of the Ethernet packet, and the payload (packet) denotes the number that is the closest to the next biggest integer. The overload is the overhead of the Ethernet packet. The value is the closest to the payload. The leader is set to 70 bytes, and the trailer is set to 66 bytes. The arrangement direction of the camera is perpendicular to the direction of the movement of the measured object. Moreover, the synchronizer simultaneously controls the starting time of each camera. The GigE camera first needs to set the host IP address and camera IP address. Besides, it decides the host network interface controller (NIC) and the address of the media access control (MAC) camera and subsequently sets the gateway and subnet mask. The system server broadcast the starting signal through the Ethernet network, and the digital signal processing (DSP) receives the signal that is leveraged to control the camera to acquire images through the general-purpose input-output (GPIO). After completion, the server broadcasts the termination signal through the Ethernet network. Besides, the DSP controls the camera to acquire the image through GPIO. The choice of light source directly influences the quality of the image acquisition and largely determines whether the defects can be revealed in the image.

**Module 2.** The choice of an embedded image processor is closely related to the real-time acquisition and processing of the defective surface images of the manufactured product. It is a bottleneck problem in a system design. It is necessary to customize a high-performing image processor since the real-time process of massive data requires fast data processing and calculation. Thus, the computer bus and parallel multicore network process jointly fulfill the task. However, this will inevitably cause the system to become complex, expensive, and difficult to maintain. Therefore, we jointly utilize the GigE camera + DSP + a server for multiprocessor processing to conduct the tasks of collection, processing, transmission, and storage of massive data.

The Texas Instruments’ TMS320DM648 (DaVinci) digital signal processor (DSP) as the core module of the embedded image processing system is chosen in the implementation of the proposed method. TMS320DM648 is based on TI’s latest TMS320C64x+ DSP core. The 64x series of the DSP chips have extremely higher frequencies (up to 900 MHz) and abundant peripheral interfaces. It is fully compatible with the C64X series of the DSP target codes. The chip integrates resources such as secondary Cache, 64-bit EMIF interface, high-precision video port, Gigabit Ethernet interface, and an integrated circuit (IIC) bus interface [1]. The system chooses TMS320DM648DSP as the core device of the embedded image processing system, and the four tailored submodules are incorporated to realize different subfunctions. Each camera in the system is connected to the DSP through the Ethernet interface to conduct parallel processing. Also, each image is transmitted to the system server through the tailored embedded system. Noticeably, it is necessary to ensure a high speed of image collections and appropriate network topology to guarantee of not encountering conflict when implementations are conducted.
Module 3. Convolutional neural network (CNN) was proposed by [5] in 1998. The advantages of CNN significantly grow when the input of the network is an image set. When the image set can be directly fed into the deep network, the complexity of the conventional visual recognition models is avoided. The process of feature extractions and data reconstruction of the CNN exhibit competitiveness in the modeling of the 2D data. For example, the network can extract image features including color, texture, shape, and image topological structure by itself.

It demonstrates exalted robustness and blaze computational effectiveness when the optical acknowledgment of the administration and the crookedness is invariant [4]. In the belles-lettres, AlexNet is the most ideal CNN cobweb workmanship that has established the authority of CNN on the data when the eyesight of the processor also excites the growth of the intricate erudition in the language notice, essential speech preserver, and reinforcement erudition. AlexNet trails a close-to-conclusion convolutional nerve plexus to end and classify resemblance shape. The fret form has a septenary couch, an intercept of the five strata of the convolutional sill, and a two-course of plentifully related courses [5]. It produces ReLU and dropout and begins data augmentation and pooling for data enlargement. For the five convolutional seams, three of them are supported on the workmanship sketch of AlexNet. Generally, AlexNet has a full of 5 convolutional belts and must fit as an unmixed banner to be built for manipulation of perception labors of many tempers [10].

We unravel the MobileNet fret by upgrading the test of AlexNet. The MobileNet can ameliorate the AlexNet under circumscribed ironmongery qualification [14]. It can lessen the number of parameters without losing precision. The usefulness of whippersnapper and jackanapes plexus can be observed instead of using the VGGNet in a single-shot multibox detector (SSD) as the fundamental convolutional nerve plexus. Figure 2 depicts the bare-bones twist makeup of MobileNet. On the other hand, Conv-Dww is a strongly separable volute construction that consists of an unmixed volume coping (depthwise layer, Dw) and a characteristic volume bed (pointwise sill, Pw). While Dw uses the intense twist coping of the 3 × 3 volume nucleus, Pw uses the inferior volume belt of the 1 × 1 twist nucleus. The effect of each volume is progressed by the batch normalization (BN) rule [14] and activates the cosine chastisement lineal one (ReLU). The BN rule settles the data allotment by coagulation of two-letter parameters, avoiding the disappearance of gradients and the setting of complex parameters.

The MobileNet convolution is calculated by

\[ C_i = (D_k + N) MD_p^2, \]  

(1)

The standard convolution is defined by

\[ C_2 = D_k^2 MNF_p^2, \]  

(2)

where \( D_k \) represents the breadth of the volute nucleus, \( M \) is an example of the numerousness of the input idol canal during volume, \( N \) depicts the content of the volute nucleus; \( DF \) is the latitude of the input appearance. Utilizing equations (1) and (2), MobileNet volume is only suffering from the summation of the flag volute. Thus, it is computationally cumbersome and is well subjugated. In a method to answer the haste requirements of the fault perception, we bind the two to rectify the computation velocity of the SSD. The MobileNet only uses 1/33 of the parameters of the ocular geometry cluster 16 (VGG-16) to affect the same assortment propriety when the computational burden is considered.

Each belt has a quantity standard and ReLU non-linearity when the drilling is in process, which mend the convergence of the MobileNet and the steadiness of birth form. The nonlinearity of ReLU mends the action of the shape correspondence. To hasten the volute trading operations, the speed of SSD is increased by going to 1.3 sets, and the nicty failure is only 0.1%. It more precisely causes the facture of the reticulation to discover a realistic age failure. The stronger characteristic of the proposed MobileNet workmanship is to disapprove the use of enumerating funds within the meshwork. It is conceded to wax the profoundness and breadth of the mesh while holding the computational side unchanged. The input has 4 ramifications, which are convolved or confederated with the percolate of the distinct gauge, and the extended form of the stitched together. Volume on manifold scales can have a quotation form of distinct scales. More shape richness degraded that the latest assortment decision is more exact. Before conducting 3 × 3 and 5 × 5 convolutions, we utilize a 1 × 1 volute to plan the diminution.

The input has 112,000 images with one or multiple defective regions manually annotated in the training of MobileNet. Then, the MobileNet is used to learn the knowledge of detecting defective regions. Finally, the proposed model can annotate the defective region regarding the output of the learned Mobile Net, when a new test image is inputted. Herein, we use the mAP to measure the accuracy of defective detection. The mAP utilizes two metrics calculated by

\[ P = \frac{TP}{(TP + FP)}, \]  

(3)

\[ R = \frac{TP}{(TP + FN)}, \]  

(4)

where \( TP \) and \( FP \) mean positive samples, respectively.

Utilizing equations (3) and (4) lead to the calculation of the mAP by

\[ mAP = \frac{1}{M} \sum_{k=1}^{M} AP(k) \]  

(5)

\[ = \frac{1}{M} \sum_{i=1}^{M} \int PRdR. \]

The pipeline of the proposed method is represented by Algorithm 1.
4. The Results of the Experiment and Its Analysis

4.1. The Comparison of the Results of the Proposed Method with Those of the Different Deep Models. We fine-tune the parameters of the proposed model concerning the well-known ImageNet [11] as a parameter tuning strategy. The optimizer uses the stochastic gradient descent algorithm (SGD) with an initial learning rate set to 0.001, momentum set to 0.9, weight decay set to 0.0005, and the batch size is fixed to 32 in the training process of the model. The newly incorporated convolutional network leverages the Xavier method for initialization. We set the number of iterations to 80,000, and the positive sample threshold is assigned to 0.4. The deep model-MobileNet training process is given as follows: we select the training set and the test sets from the VOC2007 dataset. The SSD-MobileNet model is utilized to detect the five types of targeted data as aforementioned. The optimal results are obtained through multiple iterations through training and the adjustment of the parameters. The variation of the loss function during the training phase is shown in Figure 2. As a result, the stability of the loss function decreases when the iteration number goes up in both training and test phases.

Table 2 presents the three different feature extraction networks, which are called ResNet18, VoVNet39, and ESPNetV2. The MobileNet proposed in this manuscript is employed to compare the accuracy of the defect detection. The mAP proposed by the MobileNet has a better result than do other deep networks used for feature extraction, which has 3.5% higher. This better accuracy result designates the advantage of the MobileNet in the capture of the defective regions in the proposed method. Hence, the speed of the defect detection is significantly higher than VoVNet39 and ESPNetV2, and it outperforms ResNet18 by over 2 frames per second. Therefore, as the number of epoch iterations increases, the value of the loss function gradually concurrently decreases. Finally, when the number of iterations reaches 15,000, the loss value stabilizes.

4.2. The Evaluation of the Objective Visual Quality. The evaluation of the objective visual quality is treated as the benchmark of the test of defect detection. Table 3 depicts that the performance of the proposed method is close to the standard subjective quality evaluation. The highly competitive performance of the proposed method suggests the necessity of discovering deep visual features in visual quality prediction. Herein, the architecture of the multiple deep quality evaluation is employed to engineer the deep features for both the training and test images. Afterward, we incorporate the important feature histogram-oriented gradients (HOGs) into the prediction model of the visual quality. The well-known gradient features are extracted to serve as the feature similarity measurement. To utilize these two features well, we concatenate them into a long vector row-wise. We compare the proposed method with the multiple IQA models and present the prediction results in Table 3. Thus, the employed feature concatenation strategy is found to be more robust. To verify the generalization capability of the proposed method, we subsequently carried out a visual quality evaluation method on three well-known datasets. These results have demonstrated the advantage of the proposed algorithm.
Table 4 presents our rule of Asher competitive performance compared with the well-assumed IQA algorithms. SSIM that is betokened as an organic likeness measurement between the respective and judgment images is used. However, SSIM cannot accomplish effective exploitation in Vahan likeness rank assessment due to the obstacle of the constructional queue war of the vehicles. The aid characteristic is not supportive. Table 4 shows the results of three common gradient operators, and Table 5 shows the comparative results of different methods. The experiment is conducted under unalterable parameters. The SROCC record is gained by the three slope operators on the tuning dataset. The flower auction of the slope speculator is chosen. Thus, we call Scharr speculator to calculate grade tips.

4.3. Ablation Study. Generally speaking, the proposed method is comprised of three key modules: 1. digital image acquisition, 2. hardware-based image preprocessing, and 3. MobileNet to

detect the defective regions. We verify the usefulness of each component by underweighting each of them. First, we replace our digital image acquisition module with a functionally reduced one in [12]. Then, we report the defect detection accuracy in Table 6. The accuracy of the defect detection has decreased by more than 10% when module 1 is removed. This observation clearly shows the advantage of the first module presented in this study. Afterward, we evaluate the importance of the second module when removed. Table 6 presents the increased accuracy of the defect detection over 7% on the three datasets when the preprocessing operation is used, which shows the necessity of preprocessing the images before conducting defect detection.
We also replaced the MobileNet with different deep architectures presented in Table 7. While the MobileNet reached the best performance, the requirement of running time was noticeably the smallest for the MobileNet.

5. Conclusions

This study proposes a perception-aware MobileNet-SSD method that consists of three key modules, namely, (1) digital image acquisition, (2) hardware-based image preprocessing, and (3) MobileNet architecture to detect the defective regions or surfaces of wood-based products. MobileNet extracts more necessary features and uses few inceptions attached to multiple feature maps to improve the network’s ability to distinguish different types of defects without losing precision. Besides, the speed of detecting defected products is significantly higher than those that are called VoVNet39 and ESPNetV2 and outperforms ResNet18 by over 2 frames per second.

Training the proposed model utilizing the constructed data of the surface defects helps classify and detect rough shavings, watermarks, and sand marks on the surface of wood-based panels by utilizing test data. The mAP of the network model in this study reaches over 0.88, and the detection speed is 60+ frames/s. By comparing the proposed feature detection model with the other three feature extraction networks, the proposed method has better efficiency and effectiveness and provides a reference solution for the real-time detection of the surface defects of the wood-based panels. Besides, hardware implementation makes the proposed method much more efficient than its counterparts.

| Methods     | Metrics | News  | Sport  | Movie  | Animation | Tourism | Game   |
|-------------|---------|-------|--------|--------|-----------|---------|--------|
| CBIQ        | SROCC   | 0.8546 | 0.7543 | 0.7765 | 0.8546    | 0.6577  | 0.6887 |
|             | KROCC   | 0.7821 | 0.7657 | 0.7657 | 0.7321    | 0.8324  | 0.7435 |
|             | PLCC    | 0.8544 | 0.7865 | 0.8324 | 0.9221    | 0.8776  | 0.8132 |
|             | RMSE    | 4.3341 | 8.3212 | 8.1132 | 5.4412    | 6.3432  | 5.3243 |
|             | SROCC   | 0.8546 | 0.7657 | 0.7878 | 0.8554    | 0.8546  | 0.8121 |
| LBIQ        | KROCC   | 0.7768 | 0.7435 | 0.7565 | 0.4354    | 0.7675  | 0.8435 |
|             | PLCC    | 0.8334 | 0.7878 | 0.8243 | 0.8556    | 0.8779  | 0.9554 |
|             | RMSE    | 3.4351 | 6.2132 | 6.5465 | 6.4546    | 5.3243  | 6.5645 |
|             | SROCC   | 0.8798 | 0.7987 | 0.8345 | 0.7687    | 0.7987  | 0.8445 |
| BLINDS-II   | KROCC   | 0.8445 | 0.7546 | 0.8334 | 0.6765    | 0.8554  | 0.8435 |
|             | PLCC    | 0.8768 | 0.8454 | 0.8343 | 0.7565    | 0.8667  | 0.6576 |
|             | RMSE    | 2.4354 | 3.6756 | 5.4542 | 4.5465    | 4.6575  | 6.7687 |
|             | SROCC   | 0.8546 | 0.7865 | 0.8132 | 0.8466    | 0.8556  | 0.8324 |
| DIIVINE     | KROCC   | 0.7687 | 0.7954 | 0.7556 | 0.7878    | 0.7121  | 0.7556 |
|             | PLCC    | 0.9132 | 0.8343 | 0.8657 | 0.8546    | 0.8334  | 0.8334 |
|             | RMSE    | 4.3541 | 4.1321 | 4.3213 | 6.5465    | 3.3413  | 4.6576 |
|             | SROCC   | 0.8798 | 0.7877 | 0.8567 | 0.8587    | 0.8556  | 0.8767 |
| BRISQUE     | KROCC   | 0.7657 | 0.7445 | 0.8343 | 0.7987    | 0.7886  | 0.8454 |
|             | PLCC    | 0.9343 | 0.8454 | 0.8565 | 0.8798    | 0.8445  | 0.8445 |
|             | RMSE    | 1.3545 | 3.554  | 3.5465 | 3.6574    | 4.6576  | 3.1214 |
|             | SROCC   | 0.8766 | 0.8435 | 0.8546 | 0.8798    | 0.8987  | 0.8768 |
| The proposed method | KROCC | 0.7676 | 0.7867 | 0.8324 | 0.9077 | 0.8445 | 0.0787 |
|             | PLCC    | 0.9132 | 0.8545 | 0.8554 | 0.9565    | 0.9324  | 0.8876 |
|             | RMSE    | 0.9435 | 2.3241 | 3.1323 | 2.5443    | 2.3421  | 3.4431 |

Table 5: The comparative results of the methods (repeated 10 times and averaged as final results).

| Modules | Ablation study | LIVE | TIE2008 | CSIQ |
|---------|----------------|------|---------|------|
| Module 1 | Before | 0.854 | 0.775  | 0.713 |
|         | After  | 0.657 | 0.657   | 0.656 |
| Module 2 | Before | 0.854 | 0.775  | 0.713 |
|         | After  | 0.795 | 0.687   | 0.632 |

Table 6: Comparison of the results with or without using the object detection based on modules 1 and 2.

| Modules | Ablation study | LIVE | TIE2008 | CSIQ |
|---------|----------------|------|---------|------|
|         | Before | 0.854 | 0.775  | 0.713 |
|         | After  | 0.657 | 0.657   | 0.656 |

Table 7: Comparison of the results with or without using the object detection based on module 3.
Data Availability

Data will be provided by the corresponding author with the third party’s request.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

The study was supported by the Science and Technology Project of State Grid Jiangsu Electric Power Co., Ltd. (no. J202119).

References

[1] C. Ruiqiao, “The influence and countermeasures of market orientation on the development of wood processing industry,” Forest Products Industry, vol. 57, no. 11, pp. 88-89, 2020.

[2] Q. Mingxia, “Research on the development of wood processing surplus reuse industry,” Forest Products Industry, vol. 56, no. 12, pp. 68-69, 2019.

[3] C. Liu, R. Luo, L. Chen, X. Wang, and Y. Zhou, “Surface defect recognition of wood-based panels based on region screening and segmentation and random forest,” Manufacturing Automation, vol. 40, no. 9, pp. 9–13, 2018.

[4] C. Liu, X. Wang, L. Chen, and H. Guo, “Surface defect recognition of fibreboard based on random forest,” Scientia Silvae Sinicae, vol. 54, no. 11, pp. 121–126, 2018.

[5] H. Guo, X. Wang, C. Liu, and Y. Zhou, “Research on adaptive fast threshold segmentation algorithm for surface defect detection of the wood-based panel,” Scientia Silvae Sinicae, vol. 54, no. 11, pp. 134–142, 2018.

[6] C. Longxian, C. Liu, X. Wang, and Y. Zhou, “Design, and implementation of an online synchronous image acquisition system for wood-based panels,” Manufacturing Automation, vol. 41, no. 01, pp. 68–71, 2019.

[7] Z. Sun and C. Lu, “Research status and the prospect of deep learning,” Computer Science, vol. 43, no. 02, pp. 1–8, 2016.

[8] P. Yang and W. Mengjie, “License plate recognition based on convolutional neural network,” in Science and Technology and Innovation, no. 17, pp. 72–73, 2020.

[9] T. Shiyu, A. Zhu, S. Zhang, C. Qingfeng, C. Ran, and H. Gang, “Target detection of underground personnel based on the deep convolutional neural network,” Industrial and Mining Automation, vol. 44, no. 11, pp. 32–36, 2018.

[10] Y. Cao, H. Li, and T.-b. Wang, “A review of research on target detection algorithms based on deep learning I,” Computer and Modernization, vol. 5, pp. 63–69, 2020.

[11] J. A. Sethian and A. Vladimirsky, “Fast methods for the Eikonal and related Hamilton-Jacobi equations on unstructured meshes,” Proceedings of the National Academy of Sciences, vol. 97, no. 11, pp. 5699–5703, 2000.

[12] K. Crane, C. Weischedel, and M. Wardetzky, “The heat method for distance computation,” Communications of the ACM, vol. 60, no. 11, pp. 90–99, 2017.

[13] A. Capozzoli, C. Curcio, A. Liseno, and S. Savarese, “A comparison of fast marching, fast sweeping, and fast iterative methods for the solution of the eikonal equation,” in Proceedings of the 2013 21st Telecommunications Forum Telfor (TELFOR), Belgrade, Serbia, November 2013.

[14] K. Crane, C. Weischedel, and M. Wardetzky, “Geodesics in heat,” ACM Transactions on Graphics, vol. 32, no. 5, pp. 1–11, 2013.