





Fast and Accurate Defects Detection for Additive Manufactured Parts by Multispectrum and Machine Learning

Lingbao Kong, Xing Peng, and Yao Chen

Abstract

Traditional defect detection methods for metal additive manufacturing (AM) have the problems of low detection efficiency and accuracy, while the existing machine learning detection algorithms are of poor adaptability and complex structure. To address the above problems, this article proposed an improved You Only Look Once version 3 (YOLOv3) algorithm to detect the surface defects of metal AM based on multispectrum. The weighted k -means algorithm is used to cluster the target samples to improve the matching degree between the prior frame and the feature layer. The network structure of YOLOv3 is modified by using the lightweight MobileNetv3 to replace the Darknet-53 in the original YOLOv3 algorithm. Dilated convolution and Inceptionv3 are added to improve the detection capability for surface defects. A multispectrum measuring system was also developed to obtain the AM surface data with defects for experimental verification. The results show that the detection accuracy in the test set by YOLOv3-MobileNetv3 network is 11% higher than that by the original YOLOv3 network on average. The detection accuracy for cracking defects of the three types of defects is significantly increased by 23.8%, and the detection speed is also increased by 18.2%. The experimental results show that the improved YOLOv3 algorithm realizes the end-to-end surface defect detection for metal AM with high accuracy and fast speed, which can be further applied for online defect detection.

Keywords: defect detection, machine learning, YOLOv3, multispectrum, additive manufacturing

Introduction

METAL ADDITIVE MANUFACTURING (AM) can rapidly produce complex products according to industrial design and shorten the production cycle. It also has the advantages of low energy consumption and has been widely used in advanced industrial manufacturing.¹ However, due to improper control of process parameters, instability of preparation materials, and harsh environment, defects such as porosity, balling, and cracking would appear on the surface of products by metal AM, which will lead to the fatigue strength and cause other properties problems in metal additive products.² An efficient measurement method is in demand to detect such defects for products quality control.

At present, a single-sensor image is still used to obtain the defect information of metal AM, which has insufficient in-

formation and poor anti-interference performance. Therefore, a multisensor image acquisition method is needed to obtain more defect information. At the same time, with the popularity of convolutional neural network and other deep learning methods in recent years, target detection algorithm based on deep learning has gradually replaced the traditional machine learning method to detect defects in metal AM products.³ There is a great need to investigate a high-speed and accurate end-to-end defect detection algorithm for defect detection in metal AM.

For the surface defects of metal parts, the traditional methods are realized by extracting feature factors manually and then classifying the defects,⁴ which are difficult to meet the actual production requirements in terms of detection accuracy, anti-disturbance, and real-time performance.⁵ With the development of computer technology and deep learning

Shanghai Engineering Research Center of Ultra-Precision Optical Manufacturing, Department of Optical Science and Engineering, School of Information Science and Technology, Fudan University, Shanghai, China.

Opposite page: Microscopic topography of a 3D printed part with defects. *Image credit:* SUPOE of Fudan University.

recently, the deep neural network has gradually become one of the main methods for surface detection.⁶ At present, a convolution neural network is widely used to extract features and classify surface defects of metal parts.⁷ For example, an end-to-end convolutional neural network is proposed to realize the high-precision classification of surface defects of metal parts.⁸

A neural network was used to generate a large number of unlabeled data before defect classification and solve the problem of inadequate training of defective samples.^{9,10} However, these detection methods and algorithms can only realize the surface defect classification of metal parts and cannot locate the defects with more complex features. Even though a defect detection network was used to realize the end-to-end surface defect detection and the accuracy rate was as good as 70–80%,¹¹ it is insufficient in detection speed, which cannot meet the demand of real-time detection in metal AM.

The essence of target detection is to find out all the concerned objects in the image and determine their position, size, and category information.^{12,13} The existing deep learning target detection algorithms can be divided into two categories: one is a two-stage detection algorithm represented by Faster Region with Convolutional Neural Network (R-CNN)¹⁴ and Mask R-CNN,¹⁵ and the other is a one-stage detection algorithm represented by You Only Look Once (YOLO)¹⁶ and Single Shot MultiBox Detector.¹⁷ The two-stage detection algorithm divides the detection task into two stages including the generation of the candidate region by region proposal network (RPN) and the detection of the category and location of the candidate area by the detection network. This method has higher accuracy but a slower detection speed.¹⁸ The one-stage detection algorithm does not need an RPN stage and can get the detection results directly; therefore, the detection speed is fast, but the detection accuracy is low.¹⁹

These two kinds of detection algorithms are currently widely used. There is a balance between the detection accuracy and the detection speed of YOLO. The detection accuracy mAP50 (average precision [AP] of m categories where the intersection ratio [IoU] >0.5) on the COCO (Common Objects in Context provided by the Microsoft team for image recognition) data set for the latest YOLOv3 algorithm is 55% with the input of 416×416, and the detection speed reaches 34 frames per second (fps), which can be taken as one of the most excellent detection algorithms.²⁰ However, as a one-stage algorithm, YOLOv3 has the limitation of low detection accuracy.²¹

As a result, this article proposed an improved YOLOv3 algorithm study surface defect detection for metal AM parts. Weighted K-means algorithm is used instead of the original K-means to cluster the target samples. The clustering center is first assigned and the weight parameter of each sample is set, and then, the clustering calculation is performed to avoid the influence of unbalance defect size on the clustering result and to improve the matching degree between the prior frame and the feature layer. Meanwhile, the network structure of YOLOv3 is modified, and MobileNetv3 is used as the backbone network. The cavity convolution and Inceptionv3 structure are introduced to improve the algorithm's ability for detecting surface defects. Finally, the surface defect data set of metal AM parts collected by a multichannel imaging system is used for verification. The experimental results are presented and discussed.

Fundamentals for YOLOv3 Algorithm

Detection principle

The YOLOv3 detection algorithm uses the Feature Pyramid Networks (FPN) method to extract multiple feature maps in different scales for detection separately, which improves the algorithm's ability to detect targets with different sizes.²² As shown in Figure 1, the input image is first divided into $S \times S$ grids, wherein the width and height of the grid are marked as c_x , c_y , respectively. In each grid, the relative center point coordinates are $(\sigma(t_x), \sigma(t_y))$ and the relative width is t_w and height is t_h . The final prediction frame (marked as the red rectangle) is obtained by the actual position, width, and height of the grid. The box in the dotted line is the prior box, and p_w and p_h are the width and height of the prior box, respectively. Taking the grid in the second row and second column in Figure 1 as an example, the center point position (b_x, b_y) , and width b_w and height b_h of the prediction box are obtained as follows:

$$\begin{cases} b_x = \sigma(t_x) + c_x \\ b_y = \sigma(t_y) + c_y \\ b_w = p_w e^{t_w} \\ b_h = p_h e^{t_h} \end{cases}. \quad (1)$$

The final values involved in the calculation of the loss function include the width and height of the prediction box, the position of the center point, the confidence of the prediction box, and classification information. The confidence of the prediction box refers to the IoU between the prediction box and the labeled box. The calculation formula for the IoU is shown in Equation (2):

$$\text{IoU} = \frac{\text{Box}_1 \cap \text{Box}_2}{\text{Box}_1 \cup \text{Box}_2}, \quad (2)$$

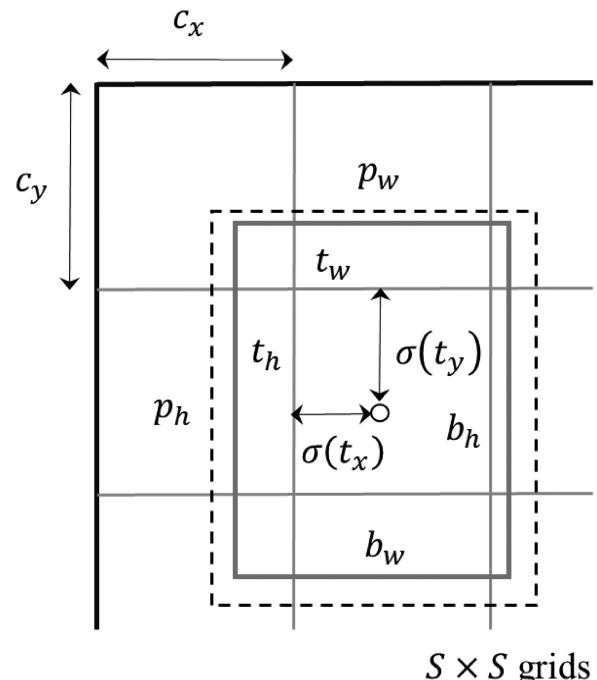


FIG. 1. Schematic diagram of prediction box.

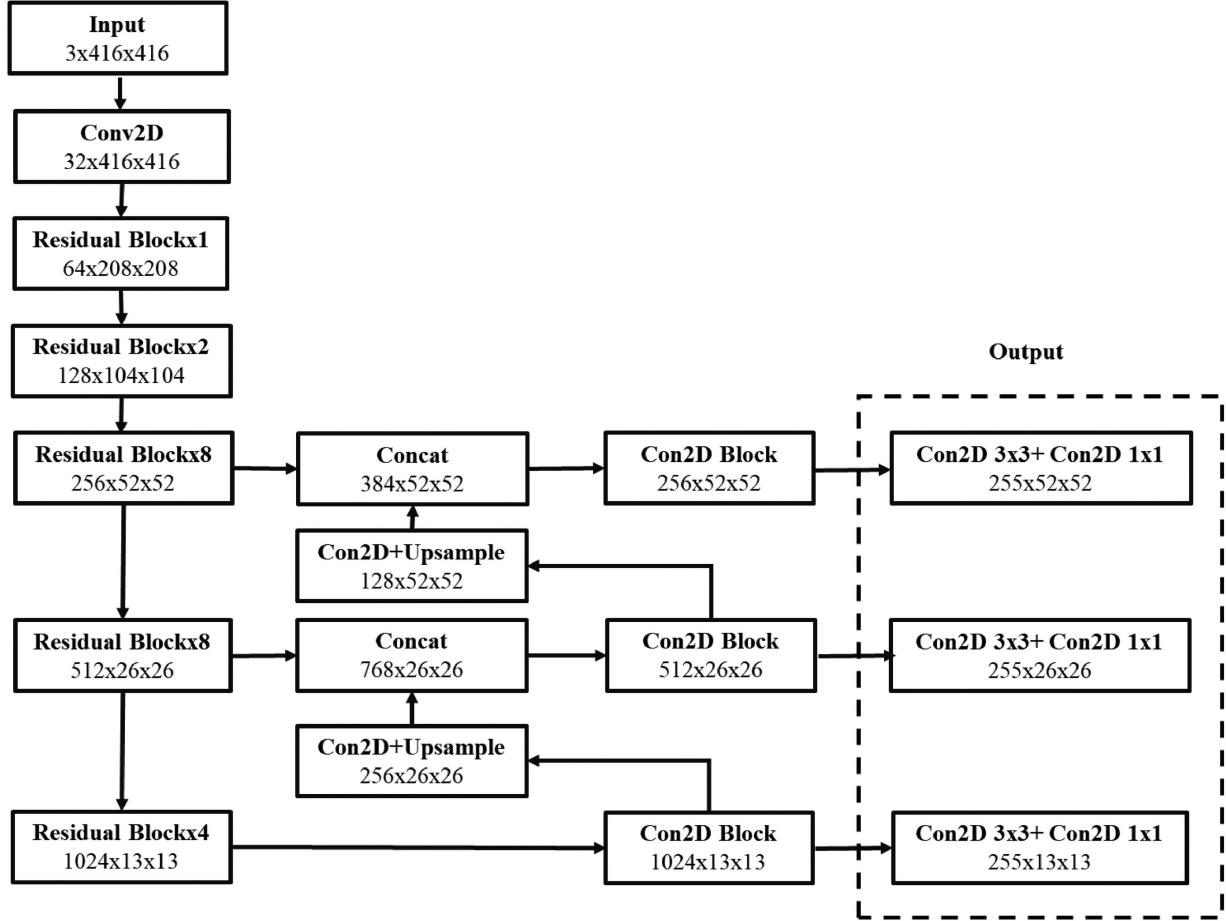


FIG. 2. Network structure diagram of YOLOv3. YOLOv3, You Only Look Once version 3.

where Box_1 is the prediction box, and Box_2 is the label box. Generally, the prediction box with $\text{IoU} > 0.7$ is set as a positive example, which means that the target is successfully predicted; while that with $\text{IoU} < 0.3$ is a negative example, which is taken as the background; and other prediction boxes will be ignored. The positive and negative examples are used to calculate the loss function. At the same time, a balance between positive and negative cases is maintained. The probability that a certain category of the target is contained in the prediction box is taken as the classification information.

Instead of a pass-through structure to detect fine-grained features in YOLOv2, YOLOv3 uses the FPN multiscale detection method combined with a residual network to convert the image into three different scale feature maps to detect large, medium, and small types of targets. The network structure of YOLOv3 is shown in Figure 2.

Loss function

Target detection requires not only the category but also the location information of the target. The two-stage detection algorithm divides the loss function into two parts: regression to the target position and classification of the target category. This approach will increase the difficulty of training and cannot achieve end-to-end detection. YOLOv3 uses only one loss function to solve the classification and location at the

same time and realize the end-to-end detection and speed up the detection algorithm.²³ The loss function is shown in Equation (3):

$$\begin{aligned} \text{Loss} = & \lambda_{coord} \sum_{i=0}^{S^2} \sum_{j=0}^B l_{ij}^{obj} [(x_i - \hat{x}_i)^2 + (y_i - \hat{y}_i)^2] \\ & + \lambda_{coord} \sum_{i=0}^{S^2} \sum_{j=0}^B l_{ij}^{obj} \omega_i [(\sqrt{\omega_i} - \sqrt{\hat{\omega}_i})^2 + (\sqrt{h_i} - \sqrt{\hat{h}_i})^2] \\ & + \sum_{i=0}^{S^2} \sum_{j=0}^B l_{ij}^{obj} (C_i - \hat{C}_i)^2 + \lambda_{noobj} \sum_{i=0}^{S^2} \sum_{j=0}^B l_{ij}^{noobj} \\ & (C_i - \hat{C}_i)^2 + \sum_{i=0}^{S^2} l_i^{noobj} \sum_{c \in classes}^B [p_i(c) - \hat{p}_i(c)]^2, \end{aligned} \quad (3)$$

where S is the image division coefficient, B is the number of prediction frames predicted in each grid, and C is the confidence parameter of classifications. In training, C_i and \hat{C}_i represent the true and the predictive value, respectively, and the value of C is determined by whether the bounding box of the grid cell is responsible for predicting an object. If responsible, then $C=1$; otherwise, $C=0$. p_i and \hat{p}_i is the true and the predictive category probability, x_i , y_i and \hat{x}_i , \hat{y}_i are the true and the predictive value of the horizontal and vertical coordinates in the i th grid, ω_i , h_i and $\hat{\omega}_i$, \hat{h}_i are the true and the predictive width and height of the center point of the box, λ_{coord} is the weight coefficient, and λ_{noobj} is the penalty weight coefficient. l_{ij}^{obj} means whether the j th anchor box of

the i th grid is responsible for this object; if it is responsible, then $l_{ij}^{obj}=1$, otherwise it is 0. l_{ij}^{noobj} means that the j th anchor box of the i th grid is not responsible for this object; if it is not responsible, then $l_{ij}^{noobj}=1$, otherwise it is 0.

Improvement of YOLOv3 Algorithm

Weighted K-means clustering algorithm

In the original YOLOv3 algorithm, multiscale features are used to detect targets, and the anchor's parameter is composed of a set of priori boxes with different sizes. The principle of using small-scale feature maps to detect large targets while using large-scale feature maps to detect small targets is also followed when assigning priori boxes. The priori boxes on feature maps of different scales should have obvious size gaps to facilitate the detection of targets with different sizes. However, the allocation of *a priori* box after the clustering completion and the corresponding relationship between the priori box and the feature map is not considered during the clustering process. Due to the imbalance of the data set, the clustering results obtained cannot match the feature map, which reduces the algorithm detection capability.

In the present study, the weighted K -means clustering algorithm is used to change the order of the clustering algorithm steps. The detection scale of each cluster center is first determined, and then, the clustering is implemented. In the clustering process, the weight of each sample is set, and it is included in the calculation of the objective function to reduce the impact of sample imbalance on the clustering result. The maximum IoU between the cluster center and the sample is used to evaluate the clustering result. The objective function D is shown in Equation (4):

$$D = \min \sum_{i=1}^n \sum_{j=1}^k \left[1 - \omega_{ijl} \frac{\text{box}_i \cap \text{cen}_j}{\text{box}_i \cup \text{cen}_j} \right], \quad (4)$$

where box_i is the actual width and height of the sample, cen_j is the j th cluster center, n is the number of samples, and k is the number of selected cluster centers. ω_{ijl} is the weight of the i th sample and the j th cluster center, and l is the center point to which the cluster center belongs.

Taking the original YOLOv3 algorithm as an example, nine cluster centers are set, and the specific implementation steps of the weighted K -means algorithm are as follows:

- (1) Setting the center points of three feature maps, and each center point corresponds to three cluster centers.
- (2) Finding the distance between each sample and the center point. The distance calculation formula is shown in Equation (5):

$$d_{il} = \sqrt{[\text{box}_i(1) - C_l(1)]^2 + [\text{box}_i(2) - C_l(2)]^2}, \quad (5)$$

where $\text{box}_i(1)$ and $\text{box}_i(2)$ are the horizontal and vertical coordinates of the i -th sample, respectively, and $C_l(1)$ and $C_l(2)$ are the horizontal and vertical coordinates of the l -th center point, respectively.

- (3) Finding the weight matrix or the weight of each sample corresponding to each cluster center. The cluster centers belonging to the same center point share the same

weight. The weight is the inverse number of the standardized distance between the sample and the cluster center, and the weight of the farther sample is directly set to 0 to reduce the interference of unbalanced samples. Equation (6) is the mean value of the distance from all samples to a certain center point, and Equation (7) is the calculation of weight.

$$\bar{d}_l = \frac{1}{n} \sum_{i=1}^n d_{il}, \quad (6)$$

$$\omega_{ijl} = \max \left(0, -\frac{d_{il} - \bar{d}_l}{\sqrt{\frac{1}{n} \sum_{i=1}^n (d_{il} - \bar{d}_l)^2}} \right), \quad (7)$$

where l is the center point, d_{il} is the distance between the sample i and the center point l , and \bar{d}_l is the average distance between all samples and the center point l .

- (4) Iteratively update all cluster centers until the cluster centers no longer change.

Network structure optimization

To make the network maintain accuracy while reducing network parameters and improve real-time performance, the following adjustments are made to the original network.

Backbone network using MobileNetv3

YOLOv3 proposes the Darknet-53 network based on the structure of ResNet-53. Although effective feature information can be extracted, the network structure is complex, which leads to too complex parameters and is not applicable in practice. Therefore, using MobileNets can effectively reduce the parameters while obtaining a certain accuracy.

MobileNetv3 is the latest version of MobileNets, which was proposed in 2019.²⁴ As a lightweight network, it has very few parameters. It mainly integrates the following features:

- (1) Combining the advantages of the previous version, the deep separable convolution of MobileNetv1 and the inverse residual structure with the linear bottleneck of MobileNetv2 are used.
- (2) The lightweight attention model based on the Squeeze and Excitation structure in MnasNet is introduced, which can better adjust the weight of each channel.
- (3) Although the swish activation function can effectively improve the accuracy of the network, upgrading the swish activation function to a hard version of swish (h-swish) can effectively improve the accuracy of the network due to a large amount of calculation of swish. The amount of calculation is largely reduced and expressed as follows:

$$h - \text{swish}(x) = x \frac{\text{Relu6}(x+3)}{6}, \quad (8)$$

where x is the independent variable of the swish activation function, Relu6 is the activation function used in MobileNetv1.

MobileNetv3 series network is used as a YOLOv3 feature extraction network. Figure 3 is the functional diagram of the MobileNetv3 convolutional network. The image is sent to the

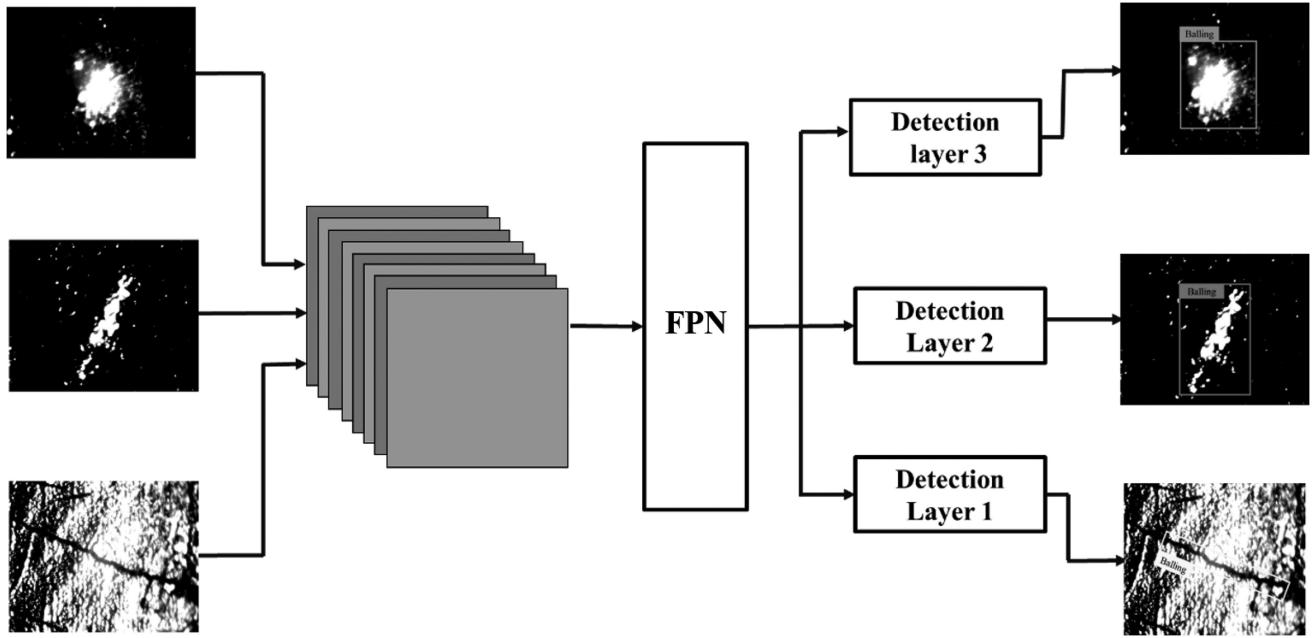


FIG. 3. MobileNetv3 convolutional network function diagram. FPN, Feature Pyramid Networks.

MobileNetv3 network and the convolutional image is sent to the FPN, as shown in Figure 3. The three detection layers detect large targets, medium targets, and small target defects, respectively. It increases the number of frames while maintaining accuracy and improves operating efficiency.

Dilated convolution is implemented by injecting holes into the standard convolution kernel to increase the receptive field of the feature map.²⁵ Figure 4a is the traditional convolution operation, and Figure 4b is the dilated convolution operation. It clearly shows that cavity convolution can not only expand the receptive field but also capture multiscale context information during convolution, which is more conducive to the detection of defects. A hyperparameter called dilation rate is added to the improved YOLOv3 network, and a double rate is added to the first feature extraction map, and a triple rate is added to the second feature extraction map. Four times rate is added to the three feature extraction images to make the image defects more obvious.

Optimization of Inceptionv3 structure

The Inceptionv3 model is the third generation of Google's Inception model.²⁶ Compared with other neural network models, the significant feature of the Inceptionv3 model is that it expands the convolution operation between neural

network layers. Inceptionv3 network uses different sizes of convolution kernel convolution so that the network layer has different sizes of perception fields and finally is stitched together to achieve different scale feature fusion. Combined with Inceptionv3, the Conv2D 3×3 convolution in the last layer of the original YOLOv3 network is changed to the Inception 3×3 convolution structure. The 1×3 convolution and 3×1 convolution are mainly used to replace the original 3×3 convolution. Then, the two 1×1 convolutions are fused to get the new Inception structure. Compared with the previous network, the number of data parameters has been greatly reduced. The structure is shown in Figure 5. Combining the above three network improvements, the overall improved network structure is shown in Figure 6.

Experimental Results and Discussion

Experimental setup

This experiment uses a multispectrum imaging system to collect the surface defect data set of metal AM. Figure 7 shows the defect measuring system, which mainly consists of a visual light/infrared camera (Type: MER-131-75GM-P NIR, pixel size: $4.8 \times 4.8 \mu\text{m}$, resolution: 1280 [H] \times 1024 [V]), objective lens (Type: St-8x-65hr, magnification: $8.0 \times$, working distance: 65 mm), and a lighting module (white light source: JC-LSD-150-W; infrared light source: JC-LSD-150-IR). To ensure the accuracy of the system, a level, a micrometer (fine adjustment stick), an XY fine-tuning disk, and a marble base are specially equipped to increase the stability. The field of view is $768 \mu\text{m} \times 614 \mu\text{m}$ and the precision is $0.6 \mu\text{m}$. White light image processing is rich in details, but it is easily disturbed by the environment.

In industrial applications, Charge Coupled Device (CCD)/Complementary Metal Oxide Semiconductor imaging devices and digital image processing technology can enhance the observation effect in complex environments. Infrared images have good thermal contrast and are less affected by

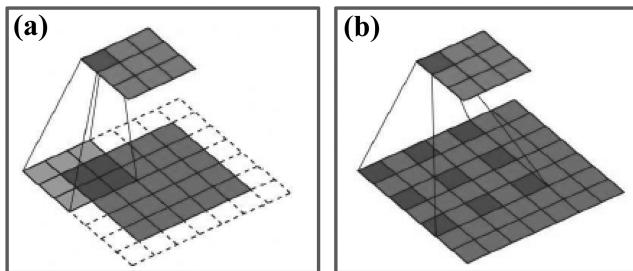


FIG. 4. Comparison of (a) traditional convolution and (b) dilated convolution operation.

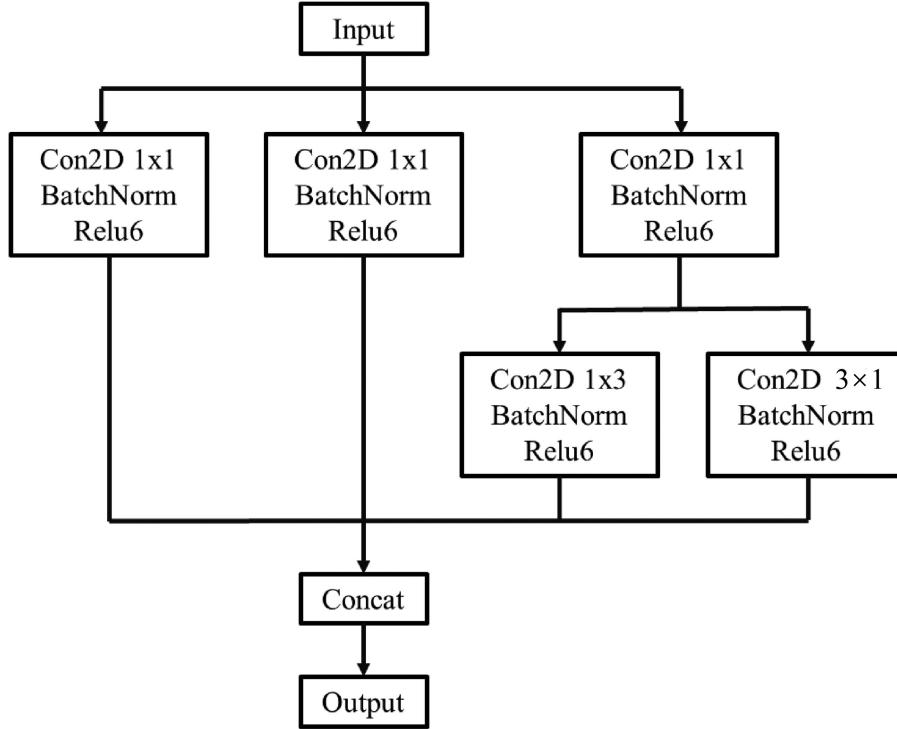


FIG. 5. Modified Inceptionv3 structure diagram.

the environment. It can effectively display the change of the thermal field during the processing. However, the infrared radiation provided by infrared imaging is very different from that of white light imaging, which affects the accurate identification of targets. During the inspection process, the light

source can be conveniently switched according to the actual detection environment. The images captured under the two detection modes of visible light and infrared have the same field of view, which is convenient for later comparison and analysis.

name	input_size	operator	exp_size	#out	SE	NL	S	output_size
The first feature extraction layers	416,416,3	conv2d		16		HS	2	208,208,16
	208,208,16	bneck,3x3	16	16		RE	1	208,208,16
	208,208,16	bneck,3x3	64	24		RE	2	104,104,24
	104,104,24	bneck,3x3	72	24		RE	1	104,104,24
	104,104,24	bneck,5x5	72	40	SE	RE	2	52,52,40
	52,52,40	bneck,5x5	120	40	SE	RE	1	52,52,40
	52,52,40	bneck,5x5	120	40	SE	RE	1	52,52,40
The second feature extraction layers	52,52,40	bneck,3x3	240	80		HS	2	26,26,80
	26,26,80	bneck,3x3	200	80		HS	1	26,26,80
	26,26,80	bneck,3x3	184	80		HS	1	26,26,80
	26,26,80	bneck,3x3	184	80	SE	HS	1	26,26,80
	26,26,80	bneck,3x3	480	112	SE	HS	1	26,26,112
	26,26,112	bneck,3x3	672	112	SE	HS	1	26,26,112
The third feature extraction layers	26,26,112	bneck,5x5	672	160	SE	HS	2	13,13,160
	13,13,160	bneck,5x5	960	160	SE	HS	1	13,13,160
	13,13,160	bneck,5x5	960	160		HS	1	13,13,160
	13,13,160	conv2d,1x1		960		HS	1	13,13,960

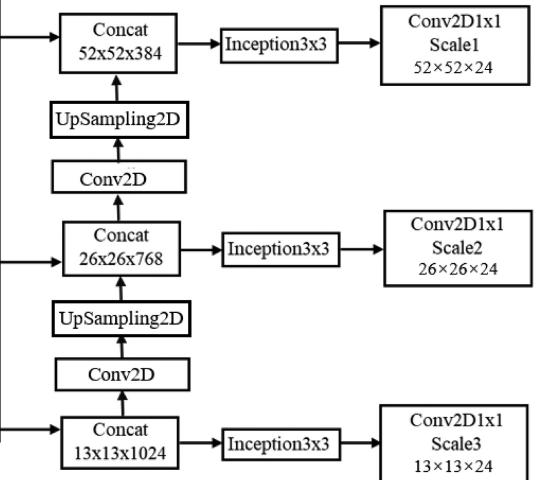


FIG. 6. The improved YOLOv3-MobileNetv3 structure diagram. HS, H-Swish; NL, the type of nonlinearity used; RE, ReLU(Linear Rectification Function); S, Stride; SE, whether there is a Squeeze-And-Excite in that block.

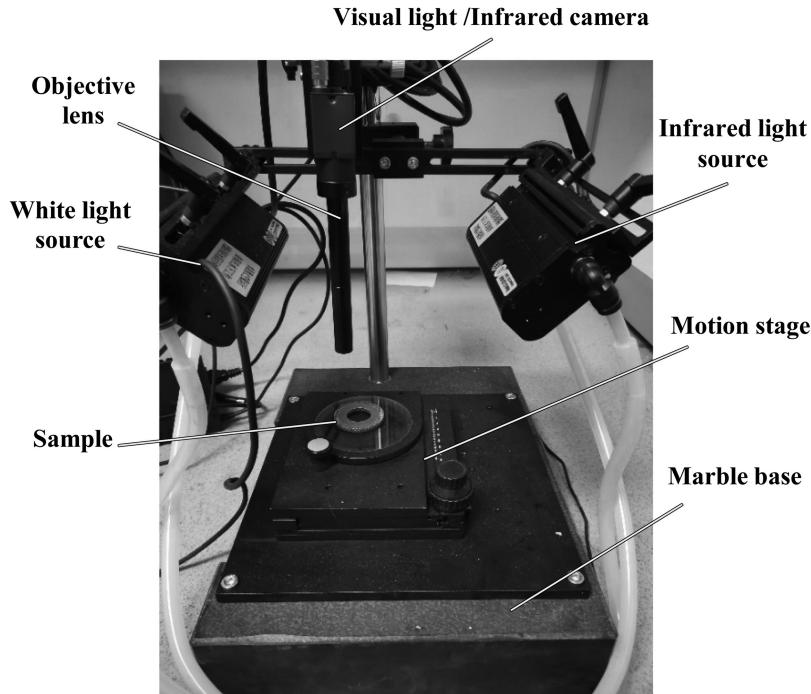


FIG. 7. Multispectrum defect measuring system.

Ti6Al4V powder was used on a commercial PBF system Arcam A2X, specifically designed to withstand extremely high process temperatures over 1100°C. The power of the electron beam can be adjusted between 50 and 3000 W. A layer thickness of 50 µm was chosen for the process with a

spot size of $\sim 250 \mu\text{m}$ in focal position. The scanning speed is 4530 mm/s and the vacuum degree is $2 \times 10^{-1} \text{ Pa}$. The observed was manufactured on a 150 mm \times 150 mm stainless steel baseplate mounted on a 200 mm \times 200 mm plate. Figure 8a illustrates the manufactured area of the specimen.

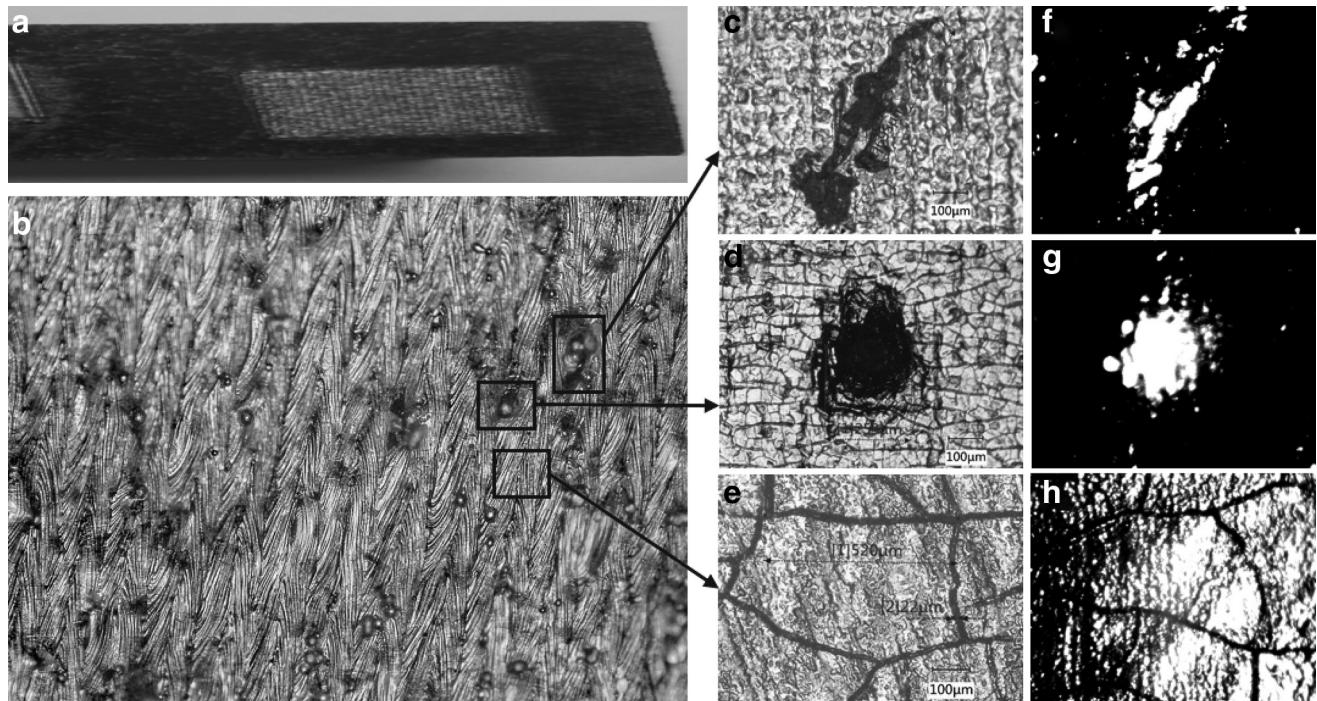


FIG. 8. The metal additive manufacturing specimen inspected by microscope and the multispectrum defect detection system. (a) Specimen. (b) The surface topography captured by the microscope in the case of a large field of view. (c–e) Porosity, balling, and cracking defects captured by the microscope, respectively. (f–h) Porosity, balling, and cracking defects captured by the developed multispectrum defect detection system, respectively.

The specimen was inspected and validation by a custom-made industrial high-accuracy 3D microscope (VHX-7000).

This microscope was equipped with 4K CMOS and high-resolution lens up to NA0.9, which facilitate that automatic observation from 20 times to 6000 times can be achieved without changing the lens. Figure 8b is captured by the microscope in the case of a large field of view. The *ex situ* inspection of the specimen confirmed the expected presence of distinctive defects. Figure 8c–e illustrates balling, porosity, and cracking defects captured by the microscope, respectively. Figure 8f–h illustrates the corresponding defects captured by the developed multispectrum defect measuring system. The images of which have been processed by image enhancement with constrained least square filtering and the top-hat algorithm.

Obviously, in the manufactured specimen, the shape of the porosity defect is irregular, and the size of the porosity is usually larger than balling, about 400–700 μm , which is usually caused by unmelted or poor melted material, as shown in Figure 8c and f. The balling defect is a regular or quasi-spherical surface defect with a size of about 200–300 μm , as presented in Figure 8d and g. The cause of balling is usually due to the mixing of air bubbles during the melting of the material. In Figure 8e and h, the width of the cracking defect is about 20 μm and the length is about 500 μm . It is usually caused by the inclusion of molten powder particles.

Totally 300 grayscale images are collected for each type of metal AM surface defect, and the size of the images is 1280 \times 1024. The ratio of the randomly selected training set and test set is 7:3 (i.e., 630 for the training set and 270 for the test set). The pictures are in XML format to record the location and category information of all defects. Meanwhile, there is no uniform division standard for the boundary of the defect, which increases the difficulty of detecting the location information.

Constrained least square filtering

For the defect image of AM, there is not enough original image data, so we choose to use constrained least-squares filtering to enhance the image. It only requires the noise variance and means value, and only these two parameters can restore the blurred image. Assuming that the input image is $f(x,y)$ and the output image is $g(x,y)$, the size of both images is $M \times N$. Each row of pixels of the image is composed of N vectors, the vector of the final single image is $MN \times 1$, and the matrix H composed of MN images has $MN \times MN$, and the output dimension of the matrix can be increased through the above change processing.

The image input and output relationship can be expressed as²⁷:

$$g(x,y) = H[f(x,y)] + \eta(x,y). \quad (9)$$

The vector expression is:

$$g = H[f(x,y)] + \eta(x,y). \quad (10)$$

After such processing, the sensitivity of the image to noise increases, and the image defects would become more and more prominent. The core of the algorithm is to study the relationship between the function H and the noise variable,

and the Laplace transform can effectively alleviate the influence of noise on the H function.

Therefore, it is expected to find the minimum criterion function C with constraints:

$$C = \sum_{x=0}^{M-1} \sum_{y=0}^{N-1} [\nabla^2 f(x,y)]^2. \quad (11)$$

The constraints are expressed as:

$$\|g - H\hat{f}\|^2 = \|\eta\|^2, \quad (12)$$

where \hat{f} is the estimation function of the original image, and ∇^2 is the Laplace operator.

The minimum value of C is required to obtain the best state of enhancing the recovery effect. Using the Lagrange method to obtain the expression of frequency-domain function C :

$$C = \|p\hat{f}\|^2, \quad (13)$$

where p is the Fourier transform of the Laplacian operator, and then, according to the Lagrange multiplier method, the expression of the Lagrange function can be obtained by substituting the function into it as:

$$\|p\hat{f}\|^2 + \lambda (\|g - H\hat{f}\|^2 - \|\eta\|^2), \quad (14)$$

where η is the Fourier transform of additive noise.

Taking the derivative of Equation (15) concerning \hat{f} , the minimum expression is obtained as:

$$\hat{f} = \frac{\lambda H^* g}{\lambda H * H + P * P} = \frac{H^* g}{H * H + \gamma P * P}, \quad (15)$$

where $p = \begin{bmatrix} 0 & -1 & 0 \\ -1 & 4 & -1 \\ 0 & -1 & 0 \end{bmatrix}$, H^* is the complex conjugate of H .

Top-hat algorithm

The top-hat algorithm takes the geometric structure of the image as the research focus and describes and analyzes the image from the perspective of collection.²⁸ Top-hat can be regarded as a mathematical operation in which grayscale image and grayscale corrosion and expansion are superimposed.

$F(x,y)$ denotes the binary function of the input image, and $S(x,y)$ denotes the definition range of the structure-function; then, the expression of the corrosion and expansion operation of S on F is as follows:

$$(F \Theta S)(x,y) = \min\{F(x-a, y-b) | (a,b) \subseteq D_S\}, \quad (16)$$

$$(F \oplus S)(x,y) = \max\{F(x+a, y+b) - S(x,b) |_{(a,b)} \subseteq D_S\}, \quad (17)$$

where the open operation and close operation of the gray image can be regarded as the superposition operation of the combination of gray corrosion and expansion.

$$F \circ S = (F \Theta S) \oplus S, \quad (18)$$

$$F \cdot S = (F \oplus S) \Theta S. \quad (19)$$

The open operation of grayscale images usually removes low-brightness image areas that are smaller than structural elements and does not affect larger low-grayscale areas. Based on the above-mentioned grayscale image-related mathematical dimension processing, the Top-hat transformation can be defined as:

$$Tophat = F - (F \circ S). \quad (20)$$

The grayscale of uneven illumination image after top-hat transformation would be appropriately corrected, which effectively improves the clarity of the defective image and makes subsequent image processing easier and more convenient.

Figure 9a is the infrared pore image taken by the multispectral defect detection system, and Figure 9b is the AM surface pore image after image preprocessing. It can be observed that the image has serious glare and the image is out of focus. After the uneven illumination of the image is processed, some defects in the image are concentrated, forming a larger patchy defect cluster, which is judged to be a pore defect with a larger defect size in AM.

Priori box parameter clustering

The priori box of the original YOLOv3 is obtained by clustering the VOC (Pascal Visual Object Classes) data set. There are significant differences in the size and type of im-

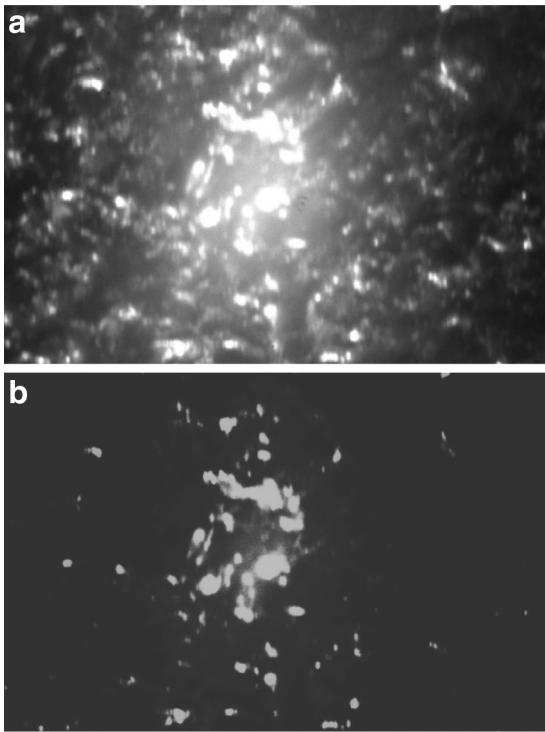


FIG. 9. Infrared image of pore defect and preprocessed image. (a) Original infrared pore image. (b) The preprocessed pore image.

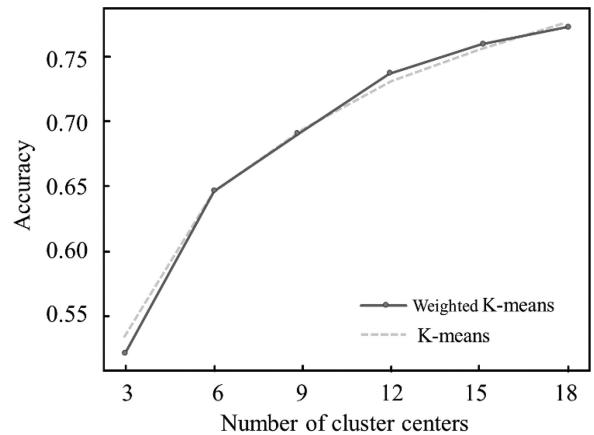


FIG. 10. Accuracy comparison for the two algorithms in different cluster centers.

ages, and it is necessary to recalculate these images. The image processing includes: first, the IoU between each sample and all cluster centers is calculated and the coordinates of the upper left corner of the sample and the cluster center are set to $(0, 0)$; second, the maximum IoU between each sample and the cluster center is taken; finally, the mean value of the calculation results is used as the accuracy of the final evaluation of the clustering results, as shown in Equation (22).

$$\text{Accuracy} = \frac{1}{n} \sum_{i=1}^n \max \left(\frac{\text{box}_i \cap \text{cen}_1}{\text{box}_i \cup \text{cen}_1}, \frac{\text{box}_i \cap \text{cen}_2}{\text{box}_i \cup \text{cen}_2}, \dots, \frac{\text{box}_i \cap \text{cen}_k}{\text{box}_i \cup \text{cen}_k} \right), \quad (21)$$

where box_i is the cluster box of each sample, cen_k is the cluster center, $\text{box}_i \cap \text{cen}_k$ means the intersection of box_i and cen_k .

Two algorithms of K -means clustering and weighted K -means clustering are used to obtain the prior frame parameters. Figure 10 shows the comparison of the two algorithms under different numbers of clustering centers. The accuracy rates of the two algorithms are calculated separately for the number of cluster centers in six cases. By calculating the accuracy of the two algorithms under different numbers of clustering centers, the results of the two algorithms are not much different, which shows that simply comparing the accuracy does not indicate the quality of the clustering.

The clustering result is expected to cover more areas, that is, the gap between the prior boxes is as large as possible. Table 1 shows the comparison of the area, width, and height of the clustering results of the two algorithms when the number of clustering centers is set to 9. After dividing the clustering results into three feature maps by size, the differences of the prior boxes between each feature map can be compared. After using the weighted K -means algorithm, the difference between the prior boxes of the three feature maps becomes larger. On the feature map of 52×52 , the area of the prior box becomes smaller, indicating that the weighted K -means can enhance the searchability of defect information, so the detection ability is stronger for defects such as balling. On the feature map of 13×13 , the area of the prior box becomes larger, indicating that the weighted K -means

TABLE 1. COMPARISON OF THE PRIOR BOX CLUSTERING RESULTS

Feature map size	K-means			Weighted K-means		
	Width	Height	Area	Width	Height	Area
52×52	31	48	1462	27	46	1184
26×26	51	117	5391	52	109	5123
13×13	124	124	16,997	116	156	17,649

can expand the location range of the defect information, so the detection ability for defects such as cracking becomes stronger.

Model training

The network construction is completed based on the improved network structure, and then, the YOLOv3-MobileNetv3 network model is experimentally trained, including a total of 800 epochs and 22 h of training (OS: Win 10; CPU: i7-9700, 16GB; GPU: RTX2070s, 8GB; Software: Python 3.6.4 [Anaconda 3.6], Tensorflow1.13.1, OpenCV4.1, Numpy1.14.2). The network is trained by the following method: first, the initial learning rate of the network is modified to 0.01, the batch size to 8, and the loss is made to drop in a large range and 300 epochs are trained to get the initial loss value; second, the learning rate is lowered to 0.001, and the loss is trained in an appropriate range for 300 epochs; finally, the learning rate is modified to 0.0001, and the loss is trained in a small range by 200 epochs to get the final training set loss value. Figure 11 shows the loss drop graph for each round of epochs of the YOLOv3-MobileNetv3 training set.

As shown in Figure 10, the network YOLOv3-MobileNetv3 drops quickly from 0 epochs to 100 epochs, and trends down slowly from 100 to 400 epochs. It stabilizes from 400 epochs on and the loss value between 5.0 and 3.0 indicates that the neural network training has a certain effect. Finally, the network converges to about 3.1.

Because YOLOv3 adopts the Darknet-53 structure, it causes a large number of parameters and insufficient memory capacity. The batch size can only be reduced to one third and modified to 4, and the training loss is 4.1. The weight file size and training time obtained are shown in Table 2.

Table 2 shows that after adding the MobileNet class structure as the backbone network and using Inceptionv3 to

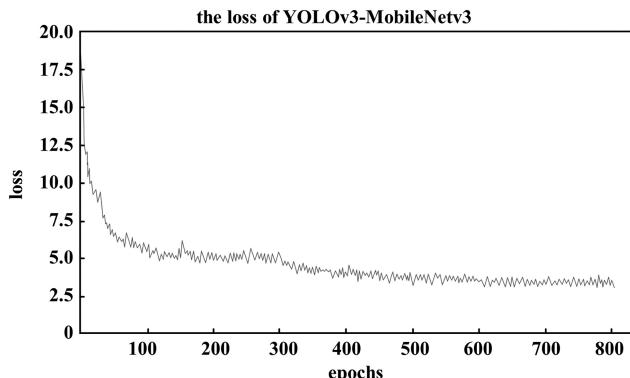


FIG. 11. Diagram of network training function change for YOLOv3-MobileNetv3.

TABLE 2. COMPARISON OF NETWORK WEIGHT FILES AND TRAINING TIME

Network structure	Weight file/MB	Training time/h
YOLOv3	235.0	8.81
YOLOv3-MobileNetv1	27.3	4.89
YOLOv3-MobileNetv2	26.6	4.75
YOLOv3-MobileNetv3	24.1	4.67

YOLOv3, You Only Look Once version 3.

reduce the total number of parameters, the weight file is reduced by about 90% compared with the original, and the training time is also shortened by about a half. A smaller weight file means that a simpler and more sophisticated network structure can be transplanted to a mobile platform with lower performance and to obtain desired results without losing accuracy.

Evaluation indexes

During training, the Adam optimizer is used to optimize network parameters. In all detections, the IoU between the actual target and the predicted target is used to evaluate whether the target position is successfully predicted. The predicted target with $\text{IoU} > 0.7$ is considered the successful prediction of the target position. When evaluating the performance of a network model, both precision (Precision) and recall (Recall) need to be taken into account. Equation (23) is the calculation formula of precision, and Equation (24) is the calculation formula of recall rate.

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}}, \quad (22)$$

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}}, \quad (23)$$

where TP is a positive example that is successfully predicted, FP is a negative example that is misjudged as a positive example by the model, and FN is a positive example that is incorrectly predicted as a negative example by the model.

AP is defined as the average value of precision under different recall rates, which is generally used to evaluate a certain type of detection accuracy. The mean average precision (mAP) of all target categories is generally used to evaluate the overall performance of the network model, as shown in Equation (25). The number of pictures that the network model can detect per second is used to evaluate the detection speed of the network model, which is defined by Vd, as shown in Equation (26).

$$\text{mAP} = \frac{\sum_{i=0}^n \text{AP}(i)}{n}, \quad (24)$$

$$\text{Vd} = \frac{\text{NumFigure}}{\text{TotalTime}}, \quad (25)$$

where $\text{AP}(i)$ is the detection accuracy of a certain category, n is the number of categories, NumFigure is the total number of detected pictures, and TotalTime is the total detection time.

The two clustering algorithms are first compared. After clustering to obtain the prior frame parameters, the original YOLOv3 algorithm is used for detection. After using the traditional K -means algorithm, the detection accuracy of different defects is quite different, and the average accuracy of porosity and balling is higher, while the average accuracy of cracking is very low, most of which cannot be detected. After using the weighted K -means algorithm, the average accuracy of most defect categories has been improved, especially for cracking, which is increased from 46.2% to 54.8%, and that of porosity is increased from 81.5% to 86.9%.

After using the weighted K -means to adjust the anchor parameters, due to the increase in the size difference between the prior frames, the detection ability for the balling type of moderately sized defects is slightly reduced, but the detection ability for the large or small defects is significantly strengthened. On the whole, the detection accuracy of the modified algorithm is higher than that before the modification; at the same time, this improvement effectively solves the problem of low detection accuracy of some defects and enhances the applicability of the algorithm.

Results and Discussion

The defects detection is implemented by the improved YOLOv3-MobileNetv3. The trained YOLOv3-MobileNetv3 was used to detect the defect images of the test set, and the resulting detection image is shown in Figure 12, in which rectangles indicate the detection labels of balling, porosity, and cracking defects. The defect features in the images are obvious, which proves that the trained model has better detection results for defects, and the algorithm in the three-feature map size can promote and feasibility for the detection of three types of defects.

Table 3 shows the performance comparison of different algorithms and different detection on the metal AM surface defect data set. RPN, ResNet50 (50-layer residual network), and YOLOv3 are currently the best performing algorithms on this data set.

As the feature map is increased from three to four layers, and the number of candidate frames increases, the detection accuracy of all types of defects has been increased compared with the original YOLOv3 algorithm. Especially for cracking defects, the detection accuracy is increased from 46.2%

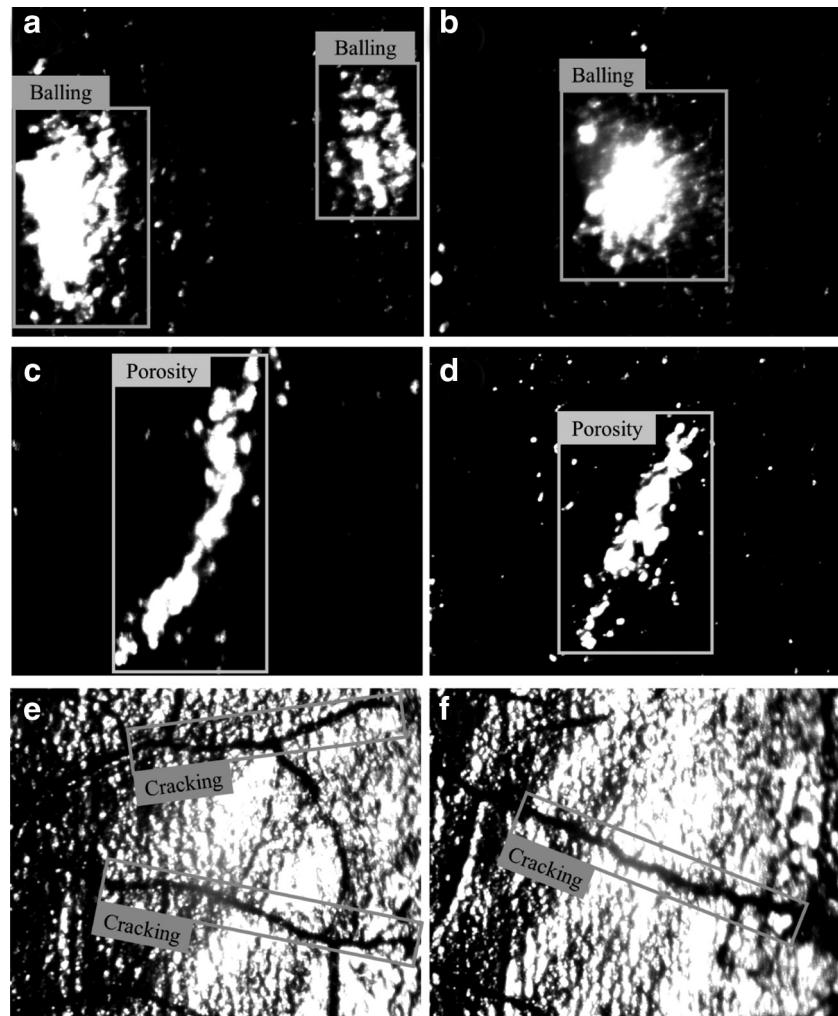


FIG. 12. Detection results of three defects (a, c, e: test images; b, d, f: detection results). (a, b) Balling test and detection result images. (c, d) Porosity test and detection result images. (e, f) Cracking test and detection result images.

TABLE 3. COMPARISON OF DEFECT DETECTION IN DIFFERENT ALGORITHMS

AP (%)	Balling	Porosity	Cracking	mAP	Speed (fps)
RPN	84.6	81.5	46.2	70.8	10
ResNet50	90.3	90.1	62.4	80.9	7
YOLOv3	84.4	87.2	44.7	72.1	55
YOLOv3-MobileNetv3	85.7	91.7	70.0	82.5	65

AP, average precision; mAP, mean average precision; RPN, region proposal network.

to 70.0%, which effectively solves the problem of low detection accuracy of this type of defect; moreover, the detection accuracy of all types of defects exceeded 70%, and the highest detection accuracy for porosity reached 91.7%. Compared with the original YOLOv3 algorithm, the improved YOLOv3-MobileNetv3 can detect all types of defects more accurately and meet the accuracy requirements for product surface defect detection in metal AM.

Comparing the detection accuracy and speed of the various algorithms on each type of defect shows that:

- (1) Compared with the ResNet50 as the feature extraction network, the improved YOLOv3-MobileNetv3 has higher detection accuracy of cracking defects, which have increased by 7.6%, but the improvement of the detection accuracy of porosity defects is not obvious and the AP of balling defects has dropped by 4.6%. Meanwhile, in the process of visual measurement image acquisition, affected by uneven light distribution and specular reflection of spheroidization defects, high-reflectivity surfaces are prone to highlight areas. The highlighted area hides the inherent information of the defect, which has a certain degree of influence on defect recognition. Meanwhile, the proposed algorithm is relatively inferior in position accuracy, and spheroidization defects may cluster in the manufactured specimen. For more concentrated defects, it is easy to cause false detection, which affects the final detection accuracy. However, the detection speed has been increased by nine times.
- (2) Compared with the original YOLOv3 algorithm, the improved YOLOv3-MobileNetv3 has improved detection accuracy on all defect categories, the mAP of three defects has increased from 72.1% to 82.5%, the fps has increased by 18%, which are a great improvement.

Therefore, the original YOLOv3 algorithm has a greater speed advantage, but its detection accuracy is lower. And the improved YOLOv3-MobileNetv3 inherits the rapidity of the original algorithm, and the detection accuracy reaches 82.5%, which makes up for the lack of accuracy of the original YOLOv3 algorithm.

Conclusions

This article proposed an improved YOLOv3 algorithm that the weighted k -means algorithm is used to cluster the target

samples to improve the matching degree between the prior frame and the feature layer. The network structure of YOLOv3 is modified by using the lightweight MobileNetv3 to replace the Darknet-53 in the original YOLOv3 algorithm. Dilated convolution and Inceptionv3 are added to improve the detection capability for surface defects. A multispectrum measuring system was also developed to obtain the AM surface data with defects for experimental verification. The results show that the priori box parameter clustering algorithm in the YOLOv3 algorithm is improved, and the weighted K -means algorithm is used to replace the traditional clustering algorithm. The obtained mAP value has increased by 4%.

The detection accuracy of all types of defects has been increased compared with the original YOLOv3 algorithm. Especially for cracking defects, the detection accuracy is increased from 46.2% to 70.0%, which effectively solves the problem of low detection accuracy of this type of defect. Moreover, the detection accuracy of all types of defects exceeded 70%, and the highest detection accuracy for porosity reached 91.7%. The network structure of the YOLOv3 algorithm is modified and optimized, and the backbone network is changed to MobileNetv3. The algorithm's ability to detect small defect samples is improved through the dilated convolution calculation and the Inceptionv3 structure.

The mAP value is increased by 10.4% compared with the original YOLOv3. The proposed YOLOv3-MobileNetv3 was verified on the metal AM surface defect data set. The mAP of surface defect detection reaches more than 70%, and the average was 82.5%. Meanwhile, the detection speed was maintained at 65 fps. The performance comparison with other algorithms shows that the improved YOLOv3-MobileNetv3 has a faster detection speed and good detection accuracy, which is better than other current deep learning surface defect detection algorithms.

Future work will focus on two aspects. One is to develop a more applicable model with a wider range of defect detection, and the other is to optimize the hardware structure for the high reflection phenomenon of some defects to improve the detection accuracy and complete the development of portable detection equipment. The MobileNet network structure can be modified, and a new data enhancement algorithm will be introduced to improve the accuracy of defect detection, so that the model can better meet the actual needs of defect detection in AM.

Author Disclosure Statement

The authors declare that there is no conflict of interests regarding this article.

Funding Information

This work was supported by the National Natural Science Foundation of China (Grant No. 52075100) and Shanghai Science and Technology Committee Innovation Grant (Grant No. 19ZR1404600).

References

1. Kong LB, Peng X, Chen Y, et al. Multi-sensor measurement and data fusion technology for manufacturing process monitoring: A literature review. Int J Extreme Manuf 2020; 2:02200.

2. Peng X, Kong LB, Chen Y, *et al.* Design of a multi-sensor monitoring system for additive manufacturing process. *Nanomanuf Metrol* 2020;3:142–150.
3. Zhu S, Du J, Ren N. A novel simple visual tracking algorithm based on hashing and deep learning. *Chin J Electron* 2017;5:189–194.
4. Xu K, Xu Y, Zhou P, *et al.* Application of RNAMlet to surface defect identification of steels. *Opt Lasers Eng* 2018; 105:110–117.
5. Song K, Yan Y. A noise robust method based on completed local binary patterns for hot-rolled steel strip surface defects. *Appl Surf Sci* 2013;285:858–864.
6. He Z, Liu Q. Deep regression neural network for industrial surface defect detection. *IEEE Access* 2020;8:35583–35591.
7. He D, Xu K, Zhou P. Defect detection of hot rolled steels with a new object detection framework called classification priority network. *Comput Ind Eng* 2018;128:290–297.
8. Fu G, Sun P, Zhu W, *et al.* A deep-learning-based approach for fast and robust steel surface defects classification. *Opt Lasers Eng* 2019;121:397–405.
9. He Y, Song K, Dong H, *et al.* Semi-supervised defect classification of steel surface based on multi-training and generative adversarial network. *Opt Lasers Eng* 2019;122: 294–302.
10. Di H, Ke X, Peng Z, *et al.* Surface defect classification of steels with a new semi-supervised learning method. *Opt Lasers Eng* 2019;117:40–48.
11. Qiu L, Wu X, Yu Z. A high-efficiency fully convolutional networks for pixel-wise surface defect detection. *IEEE Access* 2019;7:15884–15893.
12. Chen JC, Guo G, Wang WN. Artificial neural network-based online defect detection system with in-mold temperature and pressure sensors for high precision injection molding. *Int J Adv Manuf Technol* 2020;110:2023–2033.
13. Rostami J, Tse PW, Yuan M. Detection of broken wires in elevator wire ropes with ultrasonic guided waves and tone-burst wavelet. *Struct Health Monitor* 2020;19:481–494.
14. Ren S, He K, Girshick R, *et al.* Faster R-CNN: Towards real-time object detection with region proposal networks. *IEEE Trans Pattern Analysis Mach Intell* 2017;39:1137–1149.
15. Zhang L, Wu J, Fan Y, *et al.* An efficient building extraction method from high spatial resolution remote sensing images based on improved mask R-CNN. *Sensors* 2020;20: 1465.
16. Pang L, Liu H, Chen Y, *et al.* Real-time concealed object detection from passive millimeter wave images based on the YOLOv3 algorithm. *Sensors* 2020;20:1678.
17. Hu X, Li H, Li X, *et al.* MobileNet-SSD MicroScope using adaptive error correction algorithm: Real-time detection of license plates on mobile devices. *IET Intell Transp Syst* 2020;14:110–118.
18. Ren Z, Lin S, Li D, *et al.* Mask R-CNN object detection method based on improved feature pyramid. *Laser Optoelectron Prog* 2019;56:041502.
19. Nguyen DT, Nguyen TN, Kim H, *et al.* A high-throughput and power-efficient FPGA implementation of YOLO CNN for object detection. *IEEE Trans VLSI Syst* 2019;27:1861–1873.
20. Wu D, Wu Q, Yin X, *et al.* Lameness detection of dairy cows based on the YOLOv3 deep learning algorithm and a relative step size characteristic vector. *Biosyst Eng* 2020;189:150–163.
21. Wei X, Wei D, Suo D, *et al.* Multi-target defect identification for railway track line based on image processing and improved YOLOv3 model. *IEEE Access* 2020;8:61973–61988.
22. Zhang X, Dong X, Wei Q, *et al.* Real-time object detection algorithm based on improved YOLOv3. *J Electron Imaging* 2019;28:053022.1–53022.8.
23. Lyu S, Cai X, Feng R. YOLOv3 network based on improved loss function. *Comput Syst Appl* 2019;28:1–7.
24. Howard A, Sandler M, Chen B, *et al.* Searching for MobileNetV3. In: 2019 IEEE/CVF International Conference on Computer Vision (ICCV) IEEE, Seoul, Korea, October 27–November 2, 2020. IEEE.
25. Yu F, Koltun V. Multi-scale context aggregation by dilated convolutions. In: International Conference on Learning Representations (ICLR), San Juan, Puerto Rico, 2–4 May 2016.
26. Albatayneh OM, Lars Forslöf, Ksaibati K. Image retraining using TensorFlow implementation of the pre-trained Inception-v3 model for evaluating gravel road dust. *J Infrastruct Syst* 2020;26:04020014.
27. Arablouei R, Dogancay K. Linearly-constrained recursive total least-squares algorithm. *IEEE Signal Process Lett* 2012;19:821–824.
28. Deng L, Zhu H, Zhou Q, *et al.* Adaptive top-hat filter based on quantum genetic algorithm for infrared small target detection. *Multimed Tools Appl* 2018;77:10539–10551.

Address correspondence to:

Lingbao Kong
Shanghai Engineering Research Center
of Ultra-Precision Optical Manufacturing
Department of Optical Science and Engineering
School of Information Science and Technology
Fudan University
Shanghai 200438
China

E-mail: lkong@fudan.edu.cn