scientific reports

OPEN

Check for updates

A texture enhanced attention model for defect detection in thermal protection materials

Jialin Song¹, Zhaoba Wang^{1,2⊠}, Kailiang Xue¹, Youxing Chen^{1,2}, Guodong Guo¹, Maozhen Li³ & Asoke K. Nandi³

Thermal protection materials are widely used in the aerospace field, where detecting internal defects is crucial for ensuring spacecraft structural integrity and safety in extreme temperature environments. Existing detection models struggle with these materials due to challenges like defect-background similarity, tiny size, and multi-scale characteristics. Besides, there is a lack of defect datasets in real-world scenarios. To address these issues, we first construct a thermal protection material digital radiographic (DR) image dataset (TPMDR-dataset), which contains 670 images from actual production and 6,269 defect instances annotated under expert guidance. And we propose an innovative texture-enhanced attention defect detection (TADD) model that enables accurate, efficient, and real-time defect detection. To implement the TADD model, we design a texture enhancement module that can enhance the concealed defect textures and features. Then we develop a non-local dual attention module to address the issue of severe feature loss in tiny defects. Moreover, we improve the model's ability to detect multi-scale defects through a path aggregation network. The evaluation on the TPMDR-dataset and public dataset shows that the TADD model achieves a higher mean Average Precision (mAP) compared to other methods while maintaining 25 frames per second, exceeding the baseline model by 11.05%.

Keywords Defect detection, Thermal protection material, Concealed object detection, Texture enhancement, Attention mechanism

Thermal protection materials have become an essential material of selection for solid rocket motors or special thermal insulation components with the advantages of light weight, high temperature resistance, ablation resistance, and scouring resistance¹. Common thermal protection materials²include carbon fiber reinforced composites³, ceramic matrix composites⁴, and silicone rubber composites⁵. However, because of the multiple press curing between different materials is needed to form the whole, bottleneck problems such as cracking, delamination, and deformation occur frequently, which seriously affects the quality of the material. To avoid damage to the structure of special parts due to breakage of thermal protection materials, it is needed to perform efficient and real-time automatic monitoring of the structural state of the materials.

However, it is still lacking for real-time thermal protection materials defect detection. Currently, the detection methods are mainly carried out by a digital radiography (DR) system to capture the internal structure of the material^{6,7}, and then specialized and experienced operators are required to detect and judge the captured images, which is a time-consuming and labor-intensive process. Hence, it is very important to execute high-performance, fast, and automated defect detection of thermal protection materials.

Based on a large amount of labeled training data, data-driven supervised learning has played an important role in many fields such as defect recognition, fault diagnosis, and industrial manufacturing in recent years^{8–11}. Thus, it becomes crucial to construct a high-quality dataset in order to promote the development of detection methods. We construct a thermal protection material DR image dataset (TPMDR-dataset), which collects 670 DR images originated from the real scenarios of thermal protection material inspection, and we annotate 6,269 defect instances under the guidance of experienced experts, consisting of 6 defect categories, including lamination (LA), uneven layer thickness (TH), crazing (CR), internal metal inclusions (IN), transverse wrinkle (TW), and longitudinal wrinkle (LW), as illustrated in Figure 1. Lamination defects often have no distinct boundaries, and

¹School of Information and Communication Engineering, North University of China, Taiyuan 030051, China. ²National Key Laboratory of Dynamic Measurement Technology, North University of China, Taiyuan 030051, China. ³Department of Electronic and Electrical Engineering, Brunel University London, Uxbridge UB8 3PH, UK. [⊠]email: wangzb@nuc.edu.cn





present low significance appearance features; uneven layer thickness defects do not have common characteristics, and the layer can be too thick, or too thin; crazing defects are with longitudinal or oblique black dendritic lines, and the fiber texture details can be unclear; internal metal inclusions defects present an irregular distribution of dots; wrinkle defects are multiscale stripes, with different directions.

Based on the TPMDR-dataset, we construct our defect detection model by developing region-based and regression-based object detection methods. Region-based methods, such as Faster R-CNN¹², predict object categories and locations by generating proposal regions. Compared to regression-based methods, region-based methods achieve higher detection accuracy but are more time consuming. Regression-based methods, such as YOLO¹³ and SSD¹⁴, directly extract features to predict object classifications and locations, which can save a lot of time compared to region-based methods. However, to meet the detection time requirement, there are three major challenges to applying regression-based methods for defect detection in thermal protection materials:

- Defects are concealed in the background, defects and material fibers have similar texture characteristics and blurred edges between them;
- Some defects are tiny, making it difficult for the deep network to search the tiny defects, such as the metal inclusions defects shown in Figure 1(e);
- The different shapes and sizes of the same type of defects make it hard to characterize them. The above issues result in that the predicted bounding boxes could fail to accurately enclose the defect areas, causing missed detections or false positives.

To respond to the increasingly demand for high accuracy, automation, and real-time performance in thermal protection material defect detection, we develop a texture-enhanced attention detection model for thermal protection materials. To address the issue that the defect texture is similar to the background and is not easy to distinguish, we develop the feature texture enhancement module to enhance the object boundary texture features. To tackle the issue of tiny defects, we design a novel non-local dual-attention (NDA) module that combines spatial and channel information with non-local attention operations, integrating global information to enhance the model's ability to represent small targets and reduce feature loss. Finally, the path aggregation network (PAN)¹⁵ combined with NDA (named NDA-PAN) is developed to improve the robustness in detecting defects with different shapes and sizes.

In summary, the main contributions of our work are as follows:

- We construct a thermal protection material DR image dataset (TPMDR-dataset) from real-world production scenarios. It provides high-quality data in the field of thermal protection material quality monitoring, including six defect categories with a total of 670 images and 6,269 defect instances annotated under expert guidance;
- We propose a texture-enhanced attention defect detection (TADD) model for thermal protection materials, which improves the model's ability to detect concealed and tiny defects through a texture enhancement module and a non-local dual attention (NDA) module. Additionally, we develop the PAN module to enhance the model's performance in detecting defects with various shapes and sizes;
- We conduct comprehensive experiments to verify the effectiveness and robustness of our proposed method. The results show that our method achieves accurate non-destructive online defect detection at 54.74% mean average precision at 0.5 intersection over union threshold (mAP@0.5) with 25 frames per second (FPS) on TPMDR-dataset, with the capability of detecting concealed defects in complex backgrounds. Furthermore, the method is validated on the public concealed defect industrial dataset, which outperforms other related algorithms. This paper is organized as follows. Section "Related works" presents an overview of the related works. Section "Methodology" gives the details of the proposed methods. Section "Experimental results" presents extensive experiments and ablation studies. Finally, Section "Conclusion" concludes this paper.

Related works

Defect detection methods for thermal protection materials

Thermal protection materials are widely used as a composite material in aerospace field, and non-destructive testing of thermal protection materials has become very important. Traditional techniques such as microwave¹⁶, X-ray¹⁷, thermal imaging¹⁸, and terahertz imaging¹⁹ provide the essential imaging capabilities, followed by manual detection of defects. Deep learning acquires a specific feature representation of the dataset through learning from a large number of samples, which can perform precise and automated defect detection in complex material structures. Bang et al.²⁰ used the inception model to identify defects in composite materials imaging, overcoming the reliance on expert experience for detection results. Gong et al.²¹ proposed a transfer learning detection model that utilizes the domain-adaptive Faster R-CNN and feature pyramid network (FPN)²² to localize small-size void and inclusion defects in composite X-ray images. Meanwhile, Dai et al.²³in aerospace constructed an automated defect analysis system through YOLOv4²⁴ to automatically discriminate defects in solid rocket motor images under industrial computerized tomography imaging, addressing the issues of low efficiency in manual identification and data utilization. The above method provides a new solution direction for defect detection of thermal protection materials. Additionally, researchers in related fields have also explored strategies for detecting tiny defects and processing multi-scale features. Yu et al.²⁵ proposed a single-stage perceptual network to solve the problem of ineffectiveness of aggregated feature guidance module for industrial tiny defect detection. Guo et al.²⁶ proposed an improved YOLO model that incorporates a multi-scale feature fusion structure to enhance the dynamic adaptation of the detector to defects at different scales. However, there are still limitations in the application of these methods for thermal protection material detection, such as the lack of accurate detection of very tiny defects under texture interference and the lack of adaptability performance under multi-scale variations. Building on the aforementioned works, we proposed an improved model and innovative feature module design for thermal protection materials, including a non-local dual attention mechanism and an improved path aggregation network, to further optimize the detection accuracy and efficiency.

Concealed object detection method

Concealed object detection²⁷is an extensive and challenging task, aiming at identifying objects that seamlessly blend into complex backgrounds. However, existing concealed object detection methods focus on natural scenes^{27–31}, lacking adaptability to industrial environments, yet approaches in natural scenarios offer valuable insights for industrial applications.

Firstly, it is crucial to enhance the concealed object features to improve the effectiveness of concealed object detection. Liu et al.³² proposed a novel receptive field block (RFB) module inspired from the structure of the receptive field in the human visual system. The RFB takes into account the relationship between the receptive field size and the eccentricity to enhance the discriminability of features. Yuan et al.³³ proposed a receptive field block with a novel multi-feature pyramid network where asymmetric convolutional kernel is introduced in the network to enhance feature attraction by adding nonlinear transformations. Fan et al.²⁷ proposed texture enhancement module (TEM) to expand the receptive field by adding branches with larger expansion rate which enhances the concealed target features.

Secondly, neighbor connection as a strategy can effectively fuse the enhanced feature map information, which has already achieved better results in a variety of concealed object detection, e.g., Zhang et al.³¹ proposed the neighbor connection operation to enable the aggregation of feature maps from neighboring layers to reduce the computational burden and weaken the phenomenon of long term context passing. Fan et al.²⁷ proposed the neighbor connection decoder (NCD), the NCD aggregates high-level features, bridges contextual information across layers, and maintains semantic consistency within layers. In addition, Wang et al.³⁴ proposed an efficient cross-modal fusion network that utilizes RGB intensity and linear polarization cues to generate scene representations with high contrast to detect polarization-based concealed objects. Wang et al.³⁵ aggregated multilevel features and expanded the receptive field by dense concatenation, which combined the peer layer features and dual-branch features to improve the performance of concealed object detection. Liu et al.³⁶ designed the dense interactive decoder module to output rough localization maps, modulate the cross-scale feature fusion module to achieve concealed object detection.

In summary, while existing concealed object detection methods have made significant progress in natural scenes, their adaptation to industrial scenarios remains unknown. Our work addresses these gaps by incorporating an improved texture enhancement module and an adjacent decoder, specifically designed to meet the challenges of detecting concealed defects in thermal protection materials.

Methodology

Overview

Figure 2 illustrates the overall framework of our proposed TADD model, which consists of feature extraction, feature enhancement, feature fusion, and detection head modules. The features are extracted from the input image by the backbone network. Next, the feature enhancement module enhances the low saliency features and fuses them in the feature fusion module. Then the prediction bounding box is obtained by the detection head after generating anchor filters for each grid.

Furthermore, we elaborate the details of the three main modules, including (a) improved TEM, which is used to capture fine-grained textures through a larger receptive field and enhance the discriminative ability of the model to concealed defect boundaries; (b) NDA module, where non-local attention with position and channel information to synergistically optimize the prediction results of tiny targets in the deep network; (c) NDA-PAN, a bidirectional cross-scale connection to improve the network's ability to detect multi-scale targets.



Fig. 2. The structure of the proposed TADD model.

Feature extraction and enhancement

Feature Extraction. For the input image $X \in \mathbb{R}^{H \times W \times C}$ (*H*, *W*, and Crepresent the height, width, and number of channels of the input image), the output size of the three last layers of the backbone network CSPDarkNet53³⁷ is $\frac{H}{8} \times \frac{W}{8} \times 256$, $\frac{H}{16} \times \frac{W}{16} \times 384$ and $\frac{H}{32} \times \frac{W}{32} \times 512$, which covers shallow features with small receptive fields and much fine-grained feature information, and strong semantic, deep features with much global information. Significantly, the operational processing of the three-layer feature map is the key to our research.

Feature enhancement.Due to the high similarity between defect and background textures in the image, the defect features are confused with the background, which causes the model to be incapable of accurate discrimination. To address this issue, we develop an improved TEM that shares semantic features between layers via skip connections, while atrous convolution^{38,39} expands the module's receptive field, and adaptive average pooling branching is employed to address defect shape bias to generate more discriminative feature representations for concealed defects.

As shown in Figure 3, each improved TEM consists of four parallel residual branches, in each of which the first convolutional layer utilizes a 1×1 convolutional operation to reduce the channels to 32. Next, two additional layers are applied, including a convolutional layer with $(2i - 1) \times (2i - 1)$ kernels and an atrous convolutional layer with a specific dilation rate when i > 1. The parallel operation of atrous convolution with four different scale dilatation rates expands the receptive field and captures more contextual information. In particular, skip connecting the output of the previous branch. Next, the first four branches are concatenated and the channel size is recovered to C by a 1×1 convolution operation. After that, we add a parallel pooling upsampling residual branch, in which the adaptive mean pooling layer performs a global feature information extraction on the image. and then the input feature image size is recovered by a 1×1 convolution and upsampling. Finally, the whole module is fed into the ReLU function to obtain the output feature f_{output} . In addition, the depthwise separable convolutions⁴⁰ used in the module are able to reduce the computational burden without reducing the expressive power.



Fig. 3. The structure of the improved texture enhancement module.

Non-local dual attention

Deep neural networks can acquire richer and deeper semantic information through multiple convolutions on the feature maps⁴¹, but the features of tiny defects tend to weaken or even vanish during the convolution process²⁵. Meanwhile, low-level features contain more detailed information such as edges and textures, which can be used to refine the object⁴². To solve this problem, we design the NDA module to fuse low-level features with high-level features, which employs non-local attention⁴³ to handle with local and global dependencies and strengthens the model's ability to detect tiny defects in deep neural networks. It consists of two parts, Pearson product-moment correlation coefficient (PPMCC)-based position attention module and efficient channel attention module. As shown in Figure 4, the non-local dual attention for each layer fuses the input feature *x* is defined as:

$$x = f_i^{W-NCD} + f_i^{FPN} + f_{i-1}^{PAN}$$
(1)

where *i* represents the number of layers of attention insertion, f_i^{W-NCD} denotes the *i*-th layer weighted neighbour connection decoder (W-NCD) module output, f_i^{FPN} denotes the top-down FPN *i*-th layer feature branch, and f_{i-1}^{PAN} denotes the bottom-up PAN low-layer feature branch.

Weighted neighbor connection decoder

The neighbor connection decoder^{27,31} achieves more efficient learning by aggregating the enhanced three layers of features. However, it incorporates each layer's features homogeneously and does not express the current layer features saliently. Therefore, we propose the weighted neighbor connection decoder, which through weighting implements a salient representation of the current layer's features in the feature fusion process, which is part of the non-local dual attention input. As shown in Figure 5, the improved TEM enhanced three layer feature space f'_3 , f'_4 , and f'_5 is inputted into the W-NCD, and p_α , p_β , and p_γ are used as the weighting coefficients for reweighting the different output layers.



Fig. 4. The structure of the non-local dual attention.



Fig. 5. The structure of the weighted neighbour connection decoder.

PPMCC-based position attention module

The PPMCC-based positional attention module relies on the PPMCC similarity matrix multiplying with the original input features to achieve useful feature saliency and interfering feature suppression. The input feature $x \in R^{H \times W \times C}$ will be reshaped as $R^{N \times C}$, where $(N = H \times W)$ represents the total number of positions in

the feature map. Subsequently, the transpose of x is multiplied by itself to compute the dot product similarity map f. The similarity calculation in the original position attention module is defined as:

$$f(x_j, x_k) = x_j^T x_k \tag{2}$$

where *j* and *k* represent the position indices of the input feature $x \in R^{H \times W \times C}$, $x_j^T x_k$ denotes the dot product similarity. The original position attention module is subject to a softmax layer to compute the position attention map $S \in R^{N \times N}$:

$$S_{kj} = \frac{e^{f(x_j, x_k)}}{\sum_{j=1}^{N} e^{f(x_j, x_k)}}$$
(3)

 S_{kj} represents the influence of the *j*-th position on the *k*-th position. The more similar the feature representations of the two positions, the larger the numerator, indicating a higher correlation. Additionally, due to the properties of the exponential function, small changes in similarity on the feature map will be mapped to larger differences, enhancing the discrimination between targets and backgrounds.

There are two problems in the above similarity calculation process. First, when a large number of large values exist in the numerator $e^{f(x_j,x_k)}$, the denominator $\sum_{j=1}^{N} e^{f(x_j,x_k)}$ will become a very large value with respect to the numerator, resulting in the value of the mapping S_{kj} to become very small, which results in the spatial attention features that have a minimal impact on the input features. Second, when in the numerator $e^{f(x_j,x_k)}$ there are values with large differences, even if these values are very few, it will cause the positional attention mapping S_{kj} of these values to overflow and cannot participate in the subsequent gradient computation. To solve the above problem, the dot product similarity is replaced by the PPMCC, which will limit all elements $f(x_j, x_k)$ to [-1,1]. After softmax normalization, these elements are suitable to keep their differences. The PPMCC matrix is defined as:

$$f_{PPMCC}(x_j, x_k) = \frac{(x_j - \bar{x_j})^T (x_k - \bar{x_k})}{\|x_j - \bar{x_j}\| \|x_k - \bar{x_k}\|}$$
(4)

where $(x_j - \bar{x_j})^T (x_k - \bar{x_k})$ denotes the covariance of the vectors x_j and x_k to measure the degree of deviation of each dimension from its mean, $||x_j - \bar{x_j}||$ and $||x_k - \bar{x_k}||$ represent respectively the standard deviation of the vectors x_j and x_k , while $\bar{x_j}$ and $\bar{x_k}$ denote the sample mean values of vectors x_j and x_k . The PPMCC similarity map is normalized using the softmax function, and multiplied with the reshaped feature $R^{N \times C}$ to obtain the positional attention information of size $R^{H \times W \times C}$, which represents the feature information of all positions. Finally, the weighted sum of all features with the original input feature x is obtained by elementwise sum operation. As a result, the PPMCC-based positional attention module enhances the detection model's non-local contextual view and selectively aggregates contexts according to similarity map, which improves the model's ability to recognize tiny defect features, suppresses complex background features, and improves intraclass compactness and semantic consistency.

Efficient channel attention module

The channel attention module (CAM)⁴⁴exploits the interdependencies between channel maps to improve the feature representation of specific semantics, but the original channel attention module is computationally intensive and consumes more computational resources. ECA-Net⁴⁵employs adaptive selection of 1d convolutional kernel sizes to maintain performance while significantly reducing model complexity through a strategy of avoiding dimensionality reduction and interacting appropriately across channels. Inspired by⁴⁵, we integrate this attention mechanism into our NDA module, improving channel attention efficiency and aligning with our design principles.

First, a global average pooling operation is applied to the input feature $x \in R^{H \times W \times C}$ to produce a feature vector of size $R^{1 \times 1 \times C}$, which is then passed through a 1D convolution operation with a kernel size k is defined as:

$$k = \left| \frac{\log_2 C}{\gamma} + \frac{b}{\gamma} \right|_{odd} \tag{5}$$

By calculating the adaptive 1d convolution kernel size, which determines the coverage of local cross-channel interactions. Then, 1d convolution is used instead of fully connected layers to achieve cross-channel information interaction, avoiding dimensionality reduction and obtaining the weight for each channel of the feature map. It is less computationally intensive than the original channel attention, which significantly reduces the computational cost, especially for high-dimensional feature maps. Finally, the normalized weights are multiplied channel-wise with the original input feature map to generate the weighted feature map. The ECA module strengthens the inter-channel dependencies in the TADD model while maintaining feature integrity and improving the model detection efficiency.

Non-local dual attention path aggregation network

In order to obtain finer fusion features and enhance the network's robustness to scale changes. Inspired by the work of BiFPN⁴⁶, a novel NDA-PAN is designed by integrating the NDA with bidirectional cross-scale connections into the top-down and bottom-up PAN.

As shown in Figure 6, the NDA-PAN module aggregates multi-scale features at different resolutions. Formally, given a list of multiscale features (f_2 , f'_3 , f'_4 , f'_5), the goal of the NDA-PAN module is to find a transform function that can efficiently aggregate different features and output a new refined feature list. The top-down and bottom-up feature fusion processes are defined as follows:

$$P4 = Concat(up(f'_5), Conv(f'_4))$$

$$P3 = Concat(up(P4), Conv(f'_3))$$

$$P2 = Concat(up(P3), Conv(f_2))$$
(6)

$$B3 = NDA(down(P2), Conv(P3), W_NCD(f'_3))$$

$$B4 = NDA(down(B3), Conv(P4), W_NCD(f'_4))$$

$$B5 = Concat(down(B4), W_NCD(f'_5))$$
(7)

where $down(\cdot)$ denotes downsampling, $NDA(\cdot)$ denotes the operation of the proposed NDA module, and $W_NCD(\cdot)$ denotes the output of the W-NCD. Through bottom-up path enhancement, the rich texture features in the lower layers shorten the information path between the lower and topmost features. In addition, the outputs of the middle two layers (B3 and B4) are refined by the attention mechanism, which will emphasize the defect features and suppress the complex background features in the tiny defect detection task. Through the above computation, the outputs are four-scale fused features, which satisfy the detection capability of different defect sizes, improve the robustness to target scale differences, and enable all layers of NDA-PAN to share similar semantic features.

Detection layer and loss function

The detection layer is the part that further processes the output of the feature fusion to generate the predicted bounding boxes. Specifically, the detection layer adjusts the number of channels in the last dimension of NDA-PAN to $3 \times (x, y, w, h, E_{obj}, [E_{cls}^1, E_{cls}^2...E_{cls}^n])$, where 3 is the number of anchors in each grid, x, y, w, h means the coordinates and the width and height of the prediction box, E_{obj} represents the confidence score of the prediction box, the $[E_{cls}^1, E_{cls}^2...E_{cls}^n]$ stands for the probability of each category, and the size is the number of categories n.

The detection head first filters out prediction bounding boxes with confidence score E_{obj} below a preset confidence value, treating them as negative samples for initial screening. Next, it calculates the comprehensive confidence score E_{comp} for the remaining samples. The equation for calculating the comprehensive confidence score is as follows:

$$E_{comp} = E_{obj} \times max([E_{cls}^1, E_{cls}^2 \dots E_{cls}^n])$$
(8)

For prediction boxes with lower comprehensive confidence score E_{comp} , their priority is reduced in the subsequent non-maximal suppression (NMS) operation, and those with confidence below the target threshold are further filtered out. Finally, the detection head applies NMS: when the intersection over union (IOU) score *IoU* between prediction boxes exceeds the IOU threshold, it is considered as an overlap and the lower priority prediction box is removed. The final output from the detection head consists of high-confidence, non-overlapping prediction bounding boxes.



Fig. 6. Comparison of the connection between the FPN, PAN, BiFPN and NDA-PAN.

.....

The loss function consists of three components: bounding box loss, confidence loss, and classification loss, denoted as follows:

$$L_{loss} = L_{bbox} + L_{obj} + L_{cls} \tag{9}$$

The calculation of bounding box loss is performed using the method of complete intersection over union (CIoU)^{47,48}, which takes into account the area ratio, distance ratio, and aspect ratio between the predicted box and the ground truth box. CIoU accelerates model convergence and is described as follows:

$$L_{bbox} = L_{CIoU} = 1 - IoU + \frac{\rho^2(pred, gt)}{c^2(pred, gt)} + p_\alpha \cdot p_\upsilon$$
(10)

$$IoU = \frac{S_{pred} \cap S_{gt}}{S_{pred} \cup S_{gt}} \tag{11}$$

$$p_{\alpha} = \frac{p_{\upsilon}}{(1 - IoU) + p_{\upsilon}} \tag{12}$$

$$p_{\upsilon} = \frac{4}{\pi^2} \left(\arctan \frac{w_{gt}}{h_{gt}} - \arctan \frac{w_{pred}}{h_{pred}} \right)^2 \tag{13}$$

where S_{pred} represents the area of the detected defect bounding box predicted by the detector, while S_{gt} stands for the area of the manually annotated ground truth defect box. $\rho(pred, gt)$ denotes the Euclidean distance between the center points of the predicted and ground truth boxes, and c(pred, gt) represents the diagonal length of the predicted and ground truth boxes. The distance between the predicted and ground truth boxes is effectively narrowed by minimizing the ratio. p_{α} is the parameter used to do trade-off, when the IoU is larger, then p_{α} is larger, which prioritizes the aspect ratio and vice versa for the distance ratio. p_{υ} is a parameter used to measure aspect ratio consistency. w_{gt} and h_{gt} represent the width and height of the ground truth defect box, while w_{pred} and h_{pred} represent the width and height of the predicted defect box.

Both target confidence loss and classification loss employ binary cross-entropy loss. First, the confidence loss is calculated from the sample pairs obtained from positive sample matching, which includes the confidence score in the prediction box (E_{obj} as the predicted value) and the IoU value of the prediction box with the ground truth box (IoU as the ground truth). These two components are used in the binary cross-entropy loss to compute the final confidence loss. N_b is the training batch size and p_{obj} is a hyperparameter to solve the positive-negative sample imbalance issue. The equation for calculating the confidence loss is as follows:

$$L_{obj} = L_{BCE} = -\frac{1}{N_b} \sum_{k=0}^{N_b} \left(p_{obj} \cdot IoU \cdot \log(E_{obj}) + (1 - IoU) \cdot \log(1 - E_{obj}) \right)$$
(14)

Similarly, the classification loss is computed by comparing the predicted class scores of the bounding boxes with the one-hot values of the target class. The predicted class scores are passed through a sigmoid activation function before being calculated using the cross-entropy loss function. The sigmoid function maps the logits to probabilities in the range (0,1). The cross-entropy loss then measures the difference between the model's predictions and the ground truth labels, yielding the classification loss. The equation for calculating the classification loss is as follows:

$$L_{cls} = L_{CE} = -\frac{1}{N_b} \sum_{c=1}^{N_b} \sum_{c=1}^n \left(p_{cls} \cdot \log \frac{\exp(E_{cls})}{\sum_{c=1}^n \exp(E_{cls})} \cdot \hat{E}_{cls} \right)$$
(15)

where E_{cls} is the predicted class score by the model. \hat{E}_{cls} is the ground truth class value. p_{cls} is a hyperparameter to solve the category imbalance issue.

Experimental results

This section provides an experimental evaluation of the proposed TADD model. First, our assembled dataset and its details are introduced. Next, the experimental configurations and evaluation metrics are described. Third, the proposed model is compared with state-of-the-art methods. Finally, ablation studies are conducted to verify the effectiveness of the proposed module.

Dataset

Thermal protection material DR image dataset (TPMDR)

The emergence of new tasks and datasets has not only accelerated progress across various areas of computer vision but also driven significant advancements in industrial quality monitoring, as seen with datasets like PCB defect⁴⁹ and steel surface defect⁵⁰. Our assembled TMPDR-dataset aims to address specific challenges in real

		Number of	Number of defects								
TMPDR-dataset	Defective image	LA	TH	CR	IN	TW	LW				
Train set	468	1042	140	312	225	846	1755				
Validation set	67	173	17	43	40	115	298				
Test set	135	369	40	73	46	229	506				
Total	670	1584	197	428	311	1190	2559				
Average size (pixels)	4320×3556	64×173	101×375	87×227	21×24	757×248	70×647				

Table 1. TMPDR-dataset distribution.



Fig. 7. Concealed defects in industrial scenes from the CDS2K dataset.

.....

engineering applications. We collected 670 DR defect images of thermal protection materials with original resolution distributions of 4320×3556 , and classified the defects into six categories based on characteristics under the guidance of experienced experts, and carefully labeled the location information and classification of each defect, resulting in 6,269 instances in total. Table 1 shows the distribution of the TPMDR dataset for training, validation, and testing in the experiments and the average size of the defects.

Concealed defect segmentation 2K dataset (CDS2K)

As shown in Figure 7, CDS2K⁵¹ is a high-quality industrial scene concealed defect dataset, including images from pills, screws, leather, fabrics, steel surfaces, and other industrial scenes, containing a total of 1,330 defect images. These defects are selected from well-known industrial defect datasets such as MVTecAD, KolektorSDD, and CrackForest. The defects in the images of this dataset are "seamlessly" embedded in their material environment, which poses a great challenge for defect detection. Because of the limitation of the number of defects in a single scene, we chose a larger sample (the number of defects in a single scene > 50) of industrial scene images as the evaluation dataset to evaluate the detection effect of our proposed TADD model, which consists of 9 categories, 877 images, and 1,239 defect samples, with an average size of 727×712 . Table 2 shows the distribution of the original dataset scenes for the training and test data.

Experimental configurations

Experiments were conducted on a server equipped with Intel(R) Xeon(R) CPU E5-2690 v3 @ 2.60GHz and NVIDIA TU102. The pre-trained model on the COCO dataset was used to initialize the backbone network CSPDarknet53 and accelerate the convergence of the network by using a transfer learning strategy. The initial learning rate was set to 0.001, the number of iteration rounds was set to 300, and an early stopping mechanism was set to prevent overfitting. On the TMPDR-dataset, the input training image size was adjusted to 1280 × 1280 pixels, and the batch size was set to 4 due to the limitation of the GPU memory. On the CDS2K dataset, the input training image size was adjusted to 640×640 pixels, and the batch size was set to balance the importance of the three inputs in the W-NCD input. For example, on the input of the W-NCD corresponding to the B4 layer, the weight p_{β} of the input f'_4 was adjusted to 0.7, indicating that this input has a high impact on the decoder, and the rest of the inputs were weighted with $p_{\alpha} = p_{\gamma} = 0.5$ when performing the multiscale feature summation.

		Number of d	umber of defects											
CDS2K dataset	Defective image	RoadCrack	Blowhole	Break	Uneven	Pill	Screw	Zipper	Carpet	Leather				
Train set	610	230	84	92	75	79	53	126	63	69				
Validation set	83	29	11	14	9	9	8	17	8	9				
Test set	184	64	24	28	25	27	15	34	19	20				
Total	877	323	119	134	109	115	76	177	90	98				
Average size (pixels)	727×712	178×62	58×43	57×76	313×123	126×116	75×77	139×173	184×215	128×107				

Table 2. CDS2K dataset distribution.

	Weight_decay	Warmup_epochs	Warmup_momentum	Warmup_bias_lr
	0.0005	30	0.8	0.1
Training	Momentum	Anchors_list(CDS	2K dataset)	Anchors_list(TMPDR-dataset)
	0.937	[19, 19], [41, 41], [7 [133, 54], [94, 127] [206, 129], [466, 66 [159, 747], [343, 38	70, 76], , [118, 187], ;], [607, 117], ;2], [549, 297]	[6, 7], [17, 29], [19, 41], [15, 88], [21, 76], [23, 119], [19, 176], [18, 314], [25, 261], [35, 320], [219, 53], [222, 134]
Tecting	Confidence thr	eshold	NMS IoU threshold	
resung	0.001		0.5	

Table 3.	Hyperpa	arameter	during	TADD	model	training	and testin	ıg.

For each grid in the four feature maps (sizes $\frac{1}{4}$, $\frac{1}{8}$, $\frac{1}{16}$ and $\frac{1}{32}$ of the input size), we generate three different sizes of anchors by the K-means algorithm by adjusting the center coordinates, length, and width, the detection layer modifies the anchors into the final predicted bounding boxes. The a priori anchor settings for different datasets and other parameter settings are shown in Table 3:

Evaluation metrics

We select the CIoU threshold of the prediction bounding boxes to the ground truth bounding boxes larger than 0.5 as the criterion for determining target detection, and the mean average precision (mAP) as the evaluation metric. P is the precision, which represents the proportion of correctly predicted samples to all detected targets, and R is the recall, which represents the proportion of correctly predicted samples to all targets. AP is the area enclosed by the P-R curve and the coordinate axis. In addition, the number of parameters and FPS are the metrics used to evaluate the time efficiency. These metrics are defined as follows:

$$AP = \int_0^1 P(R)dR \tag{16}$$

$$mAP = \frac{\sum AP}{n_{cls}} \tag{17}$$

$$P = \frac{TP}{TP + FP} \tag{18}$$

$$R = \frac{TP}{TP + FN} \tag{19}$$

where n_{cls} is the number of sample defect categories; TP and FN are the number of defective samples predicted correctly or incorrectly, and FP represents the number of non-defective samples that are misclassified as defects.

Evaluation

Under the same environment configuration, we compare the performance of our model with the state-of-the-art object detection models, including the region-based network Faster RCNN¹², the regression-based networks baseline model (YOLOv4²⁴, CSPDarknet53 integrated with PAN), YOLO X^{52} , YOLOv8⁵³, and SSD¹⁴. We compare the performance on the TMPDR-dataset and the publicly available CDS2K dataset, separately. The results are shown in Table 4 and Table 5.

The results conducted on TMPDR are shown in Table 4. In terms of comprehensive metric mAP@0.5, our TADD model exhibits the best performance of 54.74%, which is 11.05% higher than the baseline model. Compared to Faster R-CNN, YOLO X, YOLO v8, and SSD, the mAP@0.5 of our TADD model is higher than with 8.24%, 9.25%, 3.66, and 3.81%, respectively. The higher accuracy is also obtained on the CDS2K dataset, as shown in Table 5. In terms of mAP@0.5, our TADD model also exhibits the best performance of 73.34%, which

	mAP@	0 5/%							
	- mini e	0.5770							
Detectors	LA	TH	CR	IN	TW	LW	ALL	Parameters	FPS
Faster RCNN	49.32	30.47	36.51	23.08	76.25	63.37	46.50	260.50M	6
baseline model	48.92	34.16	27.16	23.29	73.93	54.68	43.69	240.47M	28
YOLO X	50.61	23.01	29.16	23.03	81.71	65.41	45.49	202.80M	22
YOLOv8	55.68	34.47	31.27	28.49	86.22	70.34	51.08	88.26M	49
SSD	63.53	33.24	33.70	29.57	77.92	67.64	50.93	198.56M	19
TADD model	59.31	35.38	42.34	30.33	88.29	72.81	54.74	255.78M	25

Table 4. Experimental results of different detectors on TMPDR-dataset (1280×1280).

	mAP@0.5/%	nAP@0.5/%										
Detectors	.RoadCrack	Blowhole	Break	Uneven	Pill	Screw	Zipper	Carpet	Lrather	ALL	Parameters	FPS
Faster RCNN	33.74	66.71	50.82	30.70	66.25	73.29	84.71	87.28	56.55	61.12	260.50M	25
baseline model	34.77	64.61	50.92	39.15	64.26	84.84	87.47	72.28	69.81	63.12	240.47M	62
YOLO X	40.11	75.43	63.64	43.77	68.14	90.36	85.56	89.13	77.27	70.38	202.80M	39
YOLO v8	43.82	77.65	66.86	41.82	70.67	89.37	91.22	87.12	73.35	71.32	88.26M	91
SSD	36.52	60.16	52.83	42.97	59.87	83.50	87.66	71.02	69.32	62.65	198.56M	33
TADD model	40.73	78.94	67.91	42.43	72.53	93.71	94.61	93.44	75.80	73.34	255.78M	45

Table 5. Experimental results of different detectors on CDS2K dataset (640×640).

is 10.22% higher than the baseline. Compared to Faster R-CNN, YOLO X, YOLO v8, and SSD, the mAP@0.5 is higher with the amount of 12.22%, 2.96%, 2.02% and 10.69%, respectively.

The FPS results are averaged over 135 TMPDR images and 184 CDS2K images tested with the same configuration. Our TADD model achieves 25 FPS and 45 FPS, respectively, showing the fast computation in inference. In summary, our TADD model has a better detection accuracy than other detectors, and it is very fast for practical use.

Figure 8 shows the results of baseline model and our TADD model on the test images of the TMPDR-dataset. The two LW defects are closely spaced and have similar textures in group (a), the baseline model recognizes the two LW defects as one defect, while the TADD model correctly distinguishes the close textured defects. In group (b), the baseline model misses one TH defect and misclassifies one TH defect as an LA defect, while the TADD model performs correctly in the classification of TH and LA defects. In group (c), the baseline model fails to detect the hidden CR defects in TH defects, while our model does it correctly. Meanwhile, in TH defect recognition, the prediction frame formed by the TADD model wraps the defects more accurately than the baseline model. There are tiny defects IN in group (d), the TADD model correctly finds them, however, there are misses in the baseline model, which indicates that the TADD model maintains a high accuracy to cope with tiny defects. At last, on the multi-scale CR defect discrimination in group (e), our method detects more accurately and possesses a higher confidence level.

Figure 9 shows the results of baseline model and the proposed TADD model on the test images of the CDS2K dataset. It can be seen that the prediction frames generated by the proposed TADD model are the closest to the location and size distribution of the actual defects with high confidence. Moreover, The original algorithm has missed detection in group (b) and (g), and misdetection in (d), while the proposed TADD model can locate the concealed defects more accurately, as shown in (a) and (e).

Ablation studies

To evaluate the effectiveness of the improved TEM, NDA, and NDA-PAN modules in the proposed model, we conduct ablation experiments on the TMPDR and CDS2K datasets. By sequentially removing these modules from the network or replacing them with similar modules, we verify and compare the impact of each module.

Effectiveness of the improved TEM

To evaluate the effectiveness of the improved TEM module, we first removed it from the model. By comparing the results of Group 3 and Group 4 in Table 6 and Table 7, we observe a decrease in detection accuracy for the LA and CR concealed defect categories in the TMPDR-dataset, as well as for the Blowhole, Uneven, and Pill concealed defect categories in the CDS2K dataset, after the module was removed. This indicates that the improved TEM module plays a key role in addressing the issue of background texture similarity with defects.

Next, we replace our improved TEM with the texture enhancement methods proposed in³²and²⁷, inserting them at the same position in the baseline model. We then compare the module parameters, maximum receptive field, and evaluation results on the CDS2K dataset, as shown in Table 8. The method for calculating the maximum receptive field is as follows:



Fig. 8. Comparisons of detection results of our TADD model and the baseline (CSPDarkNet53+PAN) on TMPDR-dataset.



Fig. 9. Comparison of TADD model and baseline model (CSPDarkNet53+PAN) detection results and visualization of fused features by attention heat map.

	Strategy		mAP@	0.5/%								
Group	Improved TEM	W-NCD	NDA	LA	TH	CR	IN	TW	LW	ALL	Parameters	FPS
1	\checkmark	\checkmark	\checkmark	59.31	35.38	42.34	30.33	88.29	72.81	54.74	255.78M	25
2	\checkmark	\checkmark		57.11	34.47	39.22	25.36	80.56	62.50	49.87	246.78M	27
3	✓			55.81	34.88	38.96	24.51	79.04	59.95	48.86	245.06M	27
4				48.92	34.16	27.16	23.29	73.93	54.68	43.69	240.47M	28

Table 6. Effect of each component on TMPDR-dataset (1280×1280).

	Strategy	Strategy mAP@0.5/%													
Group	Improved TEM	W-NCD	NDA	RoadCrack	Blowhole	Break	Uneven	Pill	Screw	Zipper	Carpet	Leather	ALL	Parameters	FPS
1	\checkmark	\checkmark	\checkmark	40.73	78.94	67.91	42.43	72.53	93.71	94.61	93.44	75.80	73.34	255.78M	45
2	\checkmark	\checkmark		38.47	76.09	53.47	43.67	69.55	86.92	92.67	89.25	69.92	68.89	246.78M	52
3	\checkmark			37.56	75.34	51.69	42.83	71.64	86.33	92.58	87.04	69.78	68.31	245.06M	54
4				34.77	64.61	50.92	39.15	64.26	84.84	87.47	72.28	69.81	63.12	240.47M	62

Table 7. Effect of each component on CDS2K dataset (640×640) .

Module name	Parameters	Max receptive field	mAP@0.5/%
None	-	-	63.12
RFB	1,183,232	60×60	65.20(+2.08)
TEM	1,855,232	108×108	67.89(+4.77)
Improved TEM	1,577,984	200×200	68.31(+5.19)

Table 8. Performance comparison of different texture enhancement modules on baseline model.

.....

$$l_m = l_{m-1} + (f_m - 1) \times \prod_{i=1}^{m-1} S_i$$
(20)

where l_m represents the receptive field of the *m*-th layer and l_{m-1} represents the sensory field of the m-1-th layer. The initial value l_0 has a receptive field of 1, f_m represents the convolutional kernel size of the *m*-th layer, and S_i represents the step size of the *i*-th layer.

The results in Table 8 show that our improved TEM module achieves a larger maximum receptive field and a higher mean Average Precision with fewer parameters, demonstrating the superiority of our improved TEM module for the detection.

Effectiveness of the NDA module

To evaluate the effectiveness of the NDA module, we remove it in the ablation experiment. By comparing Group 1 and Group 2 in Table 6 and Table 7, we observe that after removing the NDA module, the detection accuracy of tiny defects such as IN in the TMPDR-dataset and small-sized defects like Break, Screw, and Leather in the CDS2K dataset decreased. This indicates that the NDA module effectively enhances the model's ability to detect tiny objects.

We utilize attention heat maps to compare the effects of PPMCC similarity and dot product similarity on defect features in non-local dual attention. As shown in Figure 9, In groups (c), (f), and (i), the features produced by the PPMCC similarity have a higher degree of heat mapping in the defect region, which better highlights the tiny defects and more closely resembles the actual defect distribution. In groups (b), (e), and (f), the noise in the background region is significantly suppressed compared to the dot product similarity. Overall, PPMCC similarity in non-local dual attention outperforms dot product similarity, which is more capable of highlighting tiny defect features and suppressing complex background features.

In addition, the results shown in Table 9indicate that the NDA module outperforms the dual attention network⁴⁴ under different feature fusion methods. Moreover, the number of parameters of the NDA module is less due to the integration of effective channel attention.

Effectiveness of the NDA-PAN

Finally, regarding the effectiveness of the NDA-PAN structure, we compare the impact of the proposed NDA-PAN structure and the original PAN structure. As shown in Table 9, the mAP of the NDA-PAN model is higher than that of the PAN model by 3.38% and 4.92% on the TMPDR and CDS2K datasets, respectively. This

			TMPDR-data	iset	CDS2K datas		
Feature extraction methods	Feature fusion methods	Attention mechanism	mAP@0.5%	FPS	mAP@0.5%	FPS	Parameters
	FPN(top-down)	None	47.13	31.77	60.43	77.52	158.78M
	FPN(top-down)	DAN(dot-product,CAM)	48.71	22.26	63.37	65.86	205.32M
CSPDark Nat53	FPN(top-down)	NDA(PPMCC,ECA)	50.60	25.23	65.43	69.22	169.50M
CSI Darkivet55	PAN(top-down + bottom-up)	None	49.25	26.18	63.12	50.67	240.47M
	PAN(top-down + bottom-up)	DAN(dot-product,CAM)	51.22	21.32	66.31	41.34	287.01M
	PAN(top-down + bottom-up)	NDA(PPMCC,ECA)	52.63	25.62	68.04	45.17	251.19M

Table 9. Module comparison on TMPDR-dataset and CDS2K dataset.

indicates that the proposed NDA-PAN structure can better integrate target features, resulting in higher detection accuracies.

In addition, the effectiveness of the NDA-PAN structure is evaluated by changing the structure of the feature fusion stage (e.g., FPN²², PAN¹⁵). Comparing the proposed NDA-PAN with other structures, we find that NDAPAN outperforms the other models in the table. Furthermore, the top-down + bottom-up PAN outperforms the top-down FPN under the condition of consistent attention modules. From the mAP results in the table, the top-down + bottom-up PAN has accuracies higher than the top-down FPN by 2.12%, 2.51%, and 2.03%, respectively. This verifies that bidirectional feature fusion is superior to the unidirectional approach.

Conclusion

In this paper, we have constructed a dedicated defect dataset of thermal protection materials and developed a defect detection model to address the challenges in detecting defects. We collected and annotated a thermal protection materials DR image defect detection dataset from real production scenarios, named TMPDR-dataset. This dataset contains six common yet challenging defect detection scenarios. We localized the concealed, tiny, and multi-scale characteristics of defects in the dataset to develop and verify the detection model. To enhance the texture and discrimination of concealed defects, an improved texture enhancement module has been proposed to strengthen the defect edge features. To retain tiny defect features in deeper network layers, a novel non-local dual-attention has been developed that fuses defect location and channel information to reduce feature loss and enhance the model's capture of tiny defects. Then, in order to adapt to various scale variations. Experimental results have demonstrated that the proposed model not only achieves higher detection accuracy than other models on both the TMPDR-dataset and the CDS2K dataset. The TADD model can also run at a high frame rate, achieving a processing speed of 25 FPS on an NVIDIA TU102. Our proposed detection model meets the quality monitoring requirements in real industrial production lines, providing a practical solution for DR nondestructive testing of thermal protection materials.

Data availability

The data that support the findings of this study are available from the corresponding author upon reasonable request.

Code availability

The code for our module can be found at our Github or can be requested from the corresponding author.

Received: 6 December 2024; Accepted: 5 February 2025 Published online: 10 February 2025

References

- Xue, X., Ye, Q. & Yu, Y. Thermal safety of solid rocket motor with complex charge structure. *Applied Thermal Engineering* 222, 119932. https://doi.org/10.1016/j.applthermaleng.2022.119932 (2023).
- Le, V. T., San Ha, N. & Goo, N. S. Advanced sandwich structures for thermal protection systems in hypersonic vehicles: A review. Composites Part B: Engineering 226, 109301. https://doi.org/10.1016/j.compositesb.2021.109301 (2021).
- Hou, W., Shen, Y., Jiang, K. & Wang, C. Study on mechanical properties of carbon fiber honeycomb curved sandwich structure and its application in engine hood. *Composite Structures* 286, 115302. https://doi.org/10.1016/j.compstruct.2022.115302 (2022).
- Ohnabe, H., Masaki, S., Onozuka, M., Miyahara, K. & Sasa, T. Potential application of ceramic matrix composites to aero-engine components. *Composites Part A: Applied Science and Manufacturing* 30, 489–496. https://doi.org/10.1016/S1359-835X(98)00139-0 (1999).
- Jiang, C., Jin, Y. & Gao, J. Ablation and thermal insulation properties of silicone rubber-polyarylacetylene-carbonwoven laminates for solid rocket motor. *Plastics, Rubber and Composites* 50, 362–369. https://doi.org/10.1080/14658011.2021.1909916 (2021).
- Li, L. et al. An automatic exposure imaging and enhanced display method of line scan camera for x-ray defect image of solid rocket engine shell. *Measurement* 217, 113094. https://doi.org/10.1016/j.measurement.2023.113094 (2023).
- Remakanthan, S., Kk, M., Gunasekaran, R., Thomas, C. & Thomas, C. Analysis of defects in solid rocket motors using x-ray radiography. *The E-Journal of Nondestructive Testing* 20 (2015).
- Su, B., Chen, H. & Zhou, Z. Baf-detector: An efficient cnn-based detector for photovoltaic cell defect detection. *IEEE Transactions* on Industrial Electronics 69, 3161–3171. https://doi.org/10.1109/TIE.2021.3070507 (2022).
- Qu, Z. et al. A photovoltaic cell defect detection model capable of topological knowledge extraction. Scientific Reports 14, 21904. https://doi.org/10.1038/s41598-024-72717-0 (2024).

- Shen, M. et al. Defect detection of printed circuit board assembly based on yolov5. Scientific Reports 14, 19287. https://doi.org/10. 1038/s41598-024-70176-1 (2024).
- Xu, G. et al. Research on x-ray nondestructive defect detection method of tire based on dynamic snake convolution yolo model. Scientific Reports 14, 1–18. https://doi.org/10.1038/s41598-024-80006-z (2024).
- Ren, S., He, K., Girshick, R. & Sun, J. Faster r-cnn: Towards real-time object detection with region proposal networks. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 39, 1137–1149. https://doi.org/10.1109/TPAMI.2016.2577031 (2017).
- Redmon, J., Divvala, S., Girshick, R. & Farhadi, A. You only look once: Unified, real-time object detection. In Proceedings of the IEEE conference on computer vision and pattern recognition, 779–788, https://doi.org/10.48550/arXiv.1506.02640 (2016).
- Liu, W. et al. Ssd: Single shot multibox detector. In Computer Vision-ECCV 2016: 14th European Conference, Amsterdam, The Netherlands, October 11–14, 2016, Proceedings, Part I 14, 21–37, https://doi.org/10.1007/978-3-319-46448-0_2 (Springer, 2016).
- Liu, S., Qi, L., Qin, H., Shi, J. & Jia, J. Path aggregation network for instance segmentation. In 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, 8759–8768, https://doi.org/10.1109/CVPR.2018.00913 (2018).
- Song, P. et al. Simulations and experimental study on imaging of thick defect in reusable thermal protective system using microwave ndt. *Measurement* 233, 114713. https://doi.org/10.1016/j.measurement.2024.114713 (2024).
- Bull, D., Helfen, L., Sinclair, I., Spearing, S. & Baumbach, T. A comparison of multi-scale 3d x-ray tomographic inspection techniques for assessing carbon fibre composite impact damage. *Composites Science and Technology* 75, 55–61. https://doi.org/10. 1016/j.compscitech.2012.12.006 (2013).
- Gillespie, D. I. et al. Defect detection in aerospace sandwich composite panels using conductive thermography and contact sensors. Sensors 20, 6689. https://doi.org/10.3390/s20226689 (2020).
- Xue, K. et al. Continuous terahertz wave imaging for debonding detection and visualization analysis in layered structures. *IEEE Access* 11, 31607–31618. https://doi.org/10.1109/ACCESS.2023.3252372 (2023).
- Bang, H.-T., Park, S. & Jeon, H. Defect identification in composite materials via thermography and deep learning techniques. Composite Structures 246, 112405. https://doi.org/10.1016/j.compstruct.2020.112405 (2020).
- Gong, Y., Luo, J., Shao, H. & Li, Z. A transfer learning object detection model for defects detection in x-ray images of spacecraft composite structures. *Composite Structures* 284, 115136. https://doi.org/10.1016/j.compstruct.2021.115136 (2022).
- Lin, T.-Y. et al. Feature pyramid networks for object detection. In 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 936–944, https://doi.org/10.1109/CVPR.2017.106 (2017).
- Dai, J., Li, T., Xuan, Z. & Feng, Z. Automated defect analysis system for industrial computerized tomography images of solid rocket motor grains based on yolo-v4 model. *Electronics* 11, 3215. https://doi.org/10.3390/electronics11193215 (2022).
- Bochkovskiy, A., Wang, C.-Y. & Liao, H.-Y. M. Yolov4: Optimal speed and accuracy of object detection. arXiv preprint[SPACE]arXiv:2004.10934https://doi.org/10.48550/arXiv.2004.10934 (2020).
- Yu, X., Lyu, W., Zhou, D., Wang, C. & Xu, W. Es-net: Efficient scale-aware network for tiny defect detection. *IEEE Transactions on Instrumentation and Measurement* 71, 1–14. https://doi.org/10.1109/TIM.2022.3168897 (2022).
- Guo, Z., Wang, C., Yang, G., Huang, Z. & Li, G. Msft-yolo: Improved yolov5 based on transformer for detecting defects of steel surface. Sensors 22, 3467. https://doi.org/10.3390/s22093467 (2022).
- Fan, D.-P., Ji, G.-P., Cheng, M.-M. & Shao, L. Concealed object detection. IEEE transactions on pattern analysis and machine intelligence 44, 6024–6042. https://doi.org/10.1109/TPAMI.2021.3085766 (2021).
- Ding, J., Hu, J., Lin, J. & Zhang, X. Lightweight enhanced yolov8n underwater object detection network for low light environments. Scientific Reports 14, 27922. https://doi.org/10.1038/s41598-024-79211-7 (2024).
- Ferrante, G. S., Vasconcelos Nakamura, L. H., Sampaio, S., Filho, G. P. R. & Meneguette, R. I. Evaluating yolo architectures for detecting road killed endangered brazilian animals. *Scientific reports* 14, 1353. https://doi.org/10.1038/s41598-024-52054-y (2024).
- Pang, Y., Zhao, X., Xiang, T.-Z., Zhang, L. & Lu, H. Zoomnext: A unified collaborative pyramid network for camouflaged object detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*[SPACE]https://doi.org/10.1109/TPAMI.2024.3417329 (2024).
- Zhang, C., Wang, K., Bi, H., Liu, Z. & Yang, L. Camouflaged object detection via neighbor connection and hierarchical information transfer. *Computer Vision and Image Understanding* 221, 103450. https://doi.org/10.1016/j.cviu.2022.103450 (2022).
- Liu, S., Huang, D. et al. Receptive field block net for accurate and fast object detection. In Proceedings of the European conference on computer vision (ECCV), 385–400, https://doi.org/10.48550/arXiv.1711.07767 (2018).
- Yuan, Z., Liu, Z., Zhu, C., Qi, J. & Zhao, D. Object detection in remote sensing images via multi-feature pyramid network with receptive field block. *Remote Sensing* 13, 862. https://doi.org/10.3390/rs13050862 (2021).
- Wang, X., Ding, J., Zhang, Z., Xu, J. & Gao, J. Ipnet: Polarization-based camouflaged object detection via dual-flow network. *Engineering Applications of Artificial Intelligence* 127, 107303. https://doi.org/10.1016/j.engappai.2023.107303 (2024).
- Wang, K. et al. D²c-net: A dual-branch, dual-guidance and cross-refine network for camouflaged object detection. *IEEE Transactions on Industrial Electronics* 69, 5364–5374. https://doi.org/10.1109/TIE.2021.3078379 (2022).
- Liu, Y., Li, H., Cheng, J. & Chen, X. Mscaf-net: A general framework for camouflaged object detection via learning multi-scale context-aware features. *IEEE Transactions on Circuits and Systems for Video Technology* 33, 4934–4947. https://doi.org/10.1109/T CSVT.2023.3245883 (2023).
- Wang, C.-Y. et al. Cspnet: A new backbone that can enhance learning capability of cnn. In 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), 1571–1580, https://doi.org/10.1109/CVPRW50498.2020.00203 (2020).
- Chen, L.-C., Papandreou, G., Kokkinos, I., Murphy, K. & Yuille, A. L. Deeplab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected crfs. *IEEE transactions on pattern analysis and machine intelligence* 40, 834–848. https://doi.org/10.1109/TPAMI.2017.2699184 (2017).
- Chen, L.-C., Papandreou, G., Schroff, F. & Adam, H. Rethinking atrous convolution for semantic image segmentation. arXiv preprint[SPACE]arXiv:1706.05587https://doi.org/10.48550/arXiv.1706.05587 (2017).
- Chollet, F. Xception: Deep learning with depthwise separable convolutions. In 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 1800–1807, https://doi.org/10.1109/CVPR.2017.195 (2017).
- Wu, Z., Su, L. & Huang, Q. Cascaded partial decoder for fast and accurate salient object detection. In 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 3902–3911, https://doi.org/10.1109/CVPR.2019.00403 (2019).
- 42. Zhao, T. & Wu, X. Pyramid feature attention network for saliency detection. In 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 3080–3089, https://doi.org/10.1109/CVPR.2019.00320 (2019).
- Wang, X., Girshick, R., Gupta, A. & He, K. Non-local neural networks. In 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, 7794–7803, https://doi.org/10.1109/CVPR.2018.00813 (2018).
- 44. Fu, J. et al. Dual attention network for scene segmentation. In 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 3141–3149, https://doi.org/10.1109/CVPR.2019.00326 (2019).
- 45. Wang, Q. et al. Eca-net: Efficient channel attention for deep convolutional neural networks. In 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 11531–11539, https://doi.org/10.1109/CVPR42600.2020.01155 (2020).
- Tan, M., Pang, R. & Le, Q. V. Efficientdet: Scalable and efficient object detection. In 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 10778–10787, https://doi.org/10.1109/CVPR42600.2020.01079 (2020).
- Zheng, Z. et al. Distance-iou loss: Faster and better learning for bounding box regression. In Proceedings of the AAAI conference on artificial intelligence 34, 12993–13000. https://doi.org/10.1609/aaai.v34i07.6999 (2020).
- Zheng, Z. et al. Enhancing geometric factors in model learning and inference for object detection and instance segmentation. *IEEE Transactions on Cybernetics* 52, 8574–8586. https://doi.org/10.1109/TCYB.2021.3095305 (2022).

- Ding, R., Dai, L., Li, G. & Liu, H. Tdd-net: a tiny defect detection network for printed circuit boards. CAAI Transactions on Intelligence Technology 4, 110–116. https://doi.org/10.1049/trit.2019.0019 (2019).
- 50. Song, K. & Yan, Y. A noise robust method based on completed local binary patterns for hot-rolled steel strip surface defects. Applied Surface Science 285, 858–864. https://doi.org/10.1016/j.apsusc.2013.09.002 (2013).
- 51. Fan, D.-P. et al. Advances in deep concealed scene understanding. *Visual Intelligence* 1, 16. https://doi.org/10.1007/s44267-023-00 019-6 (2023).
- Ge, Z., Liu, S., Wang, F., Li, Z. & Sun, J. Yolox: Exceeding yolo series in 2021. arXiv preprint[SPACE]arXiv:2107.08430https://doi. org/10.48550/arXiv.2107.08430 (2021).
- Jocher, G., Qiu, J. & Chaurasia, A. Ultralytics-YOLOV8. Available online: https://github.com/ultralytics/ultralytics (accessed on 1 November 2024) (2023).

Author contributions

Jialin Song: Data collection and annotation, Methodology, Programming, Experimentation and Paper writing. Zhaoba Wang: Conceptualization, Supervision and Paper reviewing. Kailiang Xue: Experimentation, Data collection and annotation. Youxing Chen: Funding acquisition and Paper Reviewing. Guodong Guo: Paper Reviewing and Editing. Maozhen Li: Paper Reviewing. Asoke K. Nandi: Paper Reviewing and Editing. The author(s) read and approved the final manuscript.

Funding

This work was supported by the Natural Science Foundation of Shanxi Province, China (202203021221118), Shanxi Scholarship Council of China, China (2022-145), and Shanxi Provincial Key Research and Development Project, China (202302020101008).

Declarations

Conflicts of interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Additional information

Correspondence and requests for materials should be addressed to Z.W.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by-nc-nd/4.0/.

© The Author(s) 2025