applied sciences

MDPI

Article

Wavelet-Enhanced YOLO for Intelligent Detection of Welding Defects in X-Ray Films

Wenyong Wu [1,2], Hongyu Cheng [2], Jiancheng Pan [2], Lili Zhong [2] and Qican Zhang [1,*]

[1] College of Electronics and Information Engineering, Sichuan University, No. 24, South Section 1, 1st Ring Road, Chengdu 610065, China; wuwenyong@stu.scu.edu.cn
[2] Chengdu Institute of Special Equipment Inspection and Testing, Chengdu 610229, China; pan799jc@zohomail.com (J.P.)
* Correspondence: zqc@scu.edu.cn

Abstract: Welding defects threaten structural integrity, demanding efficient and accurate detection methods. Traditional radiographic testing defects interpretation is subjective, necessitating automated solutions to improve accuracy and efficiency. This study integrates wavelet transform convolutions (WTConv) into YOLOv11n, creating WT-YOLO, to enhance defect detection in X-ray films. Wavelet transforms enable multi-resolution analysis, extracting both high-frequency and low-frequency features critical for detecting various welding defects. WT-YOLO replaces standard convolutional layers with WTConv, improving multi-scale feature extraction and noise suppression. Trained on 7000 radiographic images, WT-YOLO achieved a 0.0212 increase in mAP75 and a 0.0479 improvement in precision compared to YOLOv11n. On a test set of 200 images per defect category across seven defect types, WT-YOLO showed precision improvements of 0.0515 for cracks, 0.0784 for lack of fusion, 0.0067 for incomplete penetration, 0.1180 for concavity, 0.0516 for undercut, and 0.0204 for porosity, while experiencing a slight 0.0028 decline for slag inclusion. Compared to manual inspection, WT-YOLO achieved higher precision for cracks (0.0037), undercut (0.1747), slag inclusion (0.1129), and porosity (0.1074), with an inference speed 300 times faster than manual inspection. WT-YOLO enhances weld defect detection capabilities, providing the possibility for a robust solution for industrial applications.

Keywords: radiographic testing; weld defect detection; wavelet transform convolution; YOLO; multi-scale feature extraction

check for updates

Academic Editor: Antonio Di Bartolomeo

Received: 6 March 2025
Revised: 16 April 2025
Accepted: 18 April 2025
Published: 21 April 2025

Citation: Wu, W.; Cheng, H.; Pan, J.; Zhong, L.; Zhang, Q. Wavelet-Enhanced YOLO for Intelligent Detection of Welding Defects in X-Ray Films. *Appl. Sci.* **2025**, *15*, 4586. https://doi.org/10.3390/app15084586

Copyright: © 2025 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by/4.0/).

1. Introduction

Welding technology [1–3] is a metallurgical process used to join materials, creating welds with high strength, excellent sealing performance, and material continuity. This method ensures the structural integrity and lifecycle stability of products, making it indispensable across various industries. Compared to alternative joining techniques (e.g., riveting or adhesive bonding), welding is more material-efficient, reducing the need for additional components and lowering production costs. Its precision, reliability, and cost-effectiveness have led to widespread adoption in industries such as aerospace, defense, shipbuilding, chemical engineering, machinery, automotive manufacturing, and household appliances [4]. In these sectors, welding plays a critical role in producing durable and reliable joints essential for both high-performance and everyday applications.

However, due to the rapid cooling of the weld metal from a high-temperature liquid to a solid state within a short time, the welding process is prone to defects, resulting in non-uniform microstructures in the weld. Factors such as insufficient welding current,

groove contamination, unstable welding conditions, and foreign materials further exacerbate the formation of defects, including cracks, lack of fusion, incomplete penetration, concavity, undercut, slag inclusions, and porosity. These defects compromise the mechanical properties of the weld area, making it a critical vulnerability in industrial products [5]. To address these challenges, Non-Destructive Testing (NDT) techniques have been developed and widely implemented to detect and evaluate defects without damaging the material. Traditional NDT methods, including Ultrasonic Testing (UT), Radiographic Testing (RT), Magnetic particle Testing (MT), Penetrant Testing (PT), and Eddy current Testing (ET) have significantly improved the quality and efficiency of industrial products [6–9]. Recent advancements have introduced more sophisticated techniques, such as Time-of-Flight Diffraction (TOFD) ultrasonic testing [10], Phased Array Ultrasonic Testing (PAUT) [11], Computed Radiography (CR) [12], Digital Radiography (DR) [13], Acoustic Emission testing (AE) [14], and ultrasonic guided-wave testing [15]. Among these, RT stands out for its ability to penetrate materials using X-ray or γ-ray, enabling the visualization of internal structures and the detection of defects through radiation absorption and scattering. RT is widely favored for its broad applicability, strong generalizability, clear and intuitive results, ease of long-term storage, and high detection rates.

Despite its advantages, traditional RT film interpretation remains subjective, labor-intensive, and inefficient [16–18]. The manual evaluation of RT films is prone to human error and inconsistency, highlighting the need for automated defect recognition methods to enhance efficiency, standardization, and intelligence in defect detection [19–21]. Recent advancements in artificial intelligence (AI), particularly deep learning, have revolutionized the field of automated defect recognition. Convolutional neural networks (CNNs) and recurrent neural networks (RNNs) have demonstrated remarkable success in tasks such as image recognition, natural language processing, and medical diagnosis [22–25]. In the context of RT, deep learning models can automatically learn hierarchical features from raw data, eliminating the need for manual feature engineering and significantly improving accuracy and efficiency [26]. Unlike traditional methods that rely on explicit feature extraction, CNNs adopt an "end-to-end" approach, directly mapping input images to defect detection and classifications [27]. This capability has led to the development of various AI-based methods for weld seam image recognition. The deep learning-based welding defect detection process is shown in Figure 1.

In radiographic defect detection [28], methods are broadly categorized into single-stage and two-stage approaches. Single-stage methods, such as Single Shot MultiBox Detector (SSD) and You Only Look Once (YOLO), directly predict bounding box coordinates and class probabilities, offering faster inference speeds at the cost of slightly lower accuracy. In contrast, two-stage methods, such as Region-based CNNs (R-CNN, Fast R-CNN, and Faster R-CNN), first generate region proposals and then classify and refine these regions, achieving higher accuracy but with increased computational complexity. Among these, YOLO has gained significant attention for its efficiency and real-time performance. Introduced by J. Redmon et al. in 2016 [29], YOLO divides the input image into a grid and predicts bounding boxes and class probabilities for each grid cell. In recent years, YOLO has been widely applied in industrial defect detection due to its efficiency and robustness [30].

However, conventional YOLO architectures rely on standard convolutional operations, which are limited in their ability to capture multi-scale and multi-frequency features inherent in radiographic images. This limitation is particularly pronounced when detecting defects with diverse morphologies, subtle grayscale variations, and complex edge boundaries. To address this, we propose a novel approach that integrates wavelet transforms into the YOLO framework. Wavelet transforms are renowned for their multi-resolution analysis

capabilities, enabling the decomposition of signals into distinct frequency subbands [31]. By embedding learnable wavelet kernels into YOLO's backbone network, our method enhances the model's ability to dynamically extract multi-defect, multi-boundary, and multi-grayscale features.

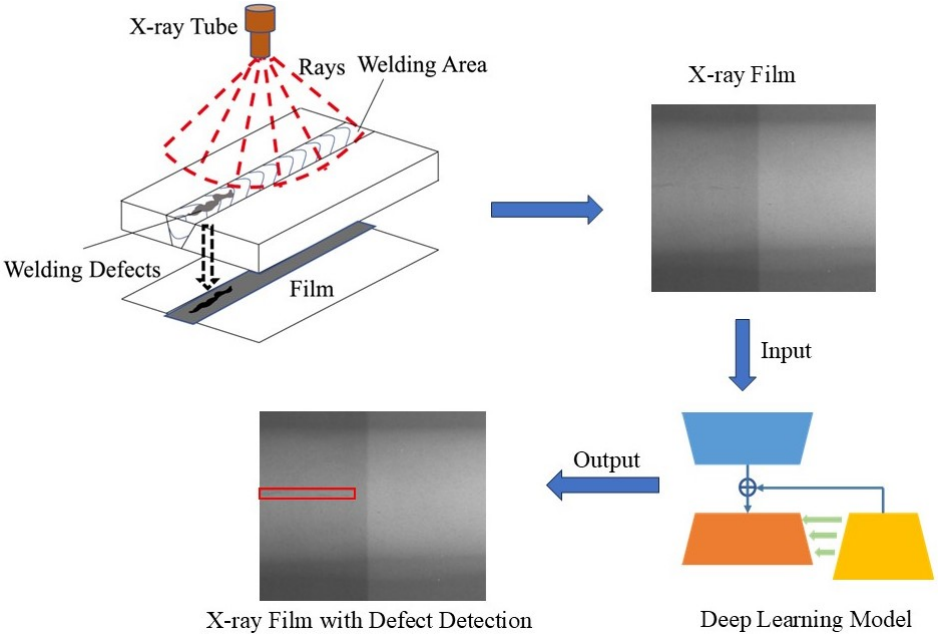


Figure 1. Deep learning-based welding defect detection process.

In this study, we leverage the multi-resolution capabilities of wavelet transforms within the YOLO framework to achieve effective multi-scale feature extraction, enabling the detection of seven defect categories: cracks, lack of fusion, incomplete penetration, concavity, undercut, slag inclusions, and porosity. Our approach represents a meaningful step forward in automated defect recognition, providing a robust and efficient solution with potential for real-world industrial applications.

Specifically, the main contributions of this paper are as follows:

(i) On a dataset composed of 7000 radiographic images, WT-YOLO achieved a 0.0212 increase in mAP75 and a 0.0479 improvement in precision compared to YOLOv11n. The wavelet-enhanced YOLO framework improves multi-scale feature extraction and defect detection accuracy.

(ii) On a test set containing seven defect types, with 200 images per type, WT-YOLO improved precision by 5.15%, 7.84%, 0.67%, 11.80%, 5.16%, and 2.04% for cracks, lack of fusion, incomplete penetration, concave, undercut, and porosity, respectively. WT-Conv's frequency decomposition effectively suppresses noise, enhancing robustness in complex industrial environments.

(iii) Compared to manual inspection, WT-YOLO achieved higher precision by 0.37%, 17.47%, 11.29%, and 10.74% for cracks, undercut, slag inclusion, and porosity, respectively, with an inference speed 300 times faster than manual inspection. The comparison between the model's performance and manual inspection results, along with the detection efficiency, provides practical support for the development of hybrid detection systems.

2. Related Work

2.1. YOLO Architectures in Radiographic Defect Detection

YOLO is a deep learning-based object detection algorithm that transforms the object detection problem into a regression task, enabling direct prediction of object locations and classes through a single neural network. The overall architecture of YOLO includes convolutional layers, fully connected layers, and an output layer. The convolutional layers extract image features, while the fully connected layers map these features to object positions and classes. The output layer provides the final predictions of object locations and classes. YOLO divides the input image into a grid of cells, with each cell responsible for detecting objects whose center points fall within it. For each cell, YOLO predicts bounding boxes along with their confidence scores and class probabilities. The Non-Maximum Suppression (NMS) algorithm is then applied to remove redundant bounding boxes, retaining those most likely to contain the object. This approach allows YOLO to perform real-time object detection with high accuracy and efficiency.

Since its inception in 2016, YOLO has evolved through multiple iterations (YOLOv1–YOLOv11) [32,33], each introducing significant innovations. For example, YOLOv2 introduced anchor boxes to improve localization accuracy [34], while YOLOv3 incorporated feature pyramid networks (FPNs) to enhance multi-scale detection capabilities [35]. Recent advancements have introduced novel components such as the C3k2 and C2PSA blocks, enhancing YOLO's feature extraction and processing capabilities. These improvements increase computational efficiency without compromising accuracy, while significantly reducing the model's parameter count [33].

In the field of radiographic defect detection, several studies have adapted YOLO to address the unique challenges posed by welding defects. For instance, Cheng et al. [36] introduced a lightweight version of YOLOv5 by simplifying the network structure, reducing the number of parameters, and optimizing computational efficiency. This improved model demonstrated higher detection accuracy and faster inference speeds when applied to X-ray images of welding defects, including porosity, slag inclusion, and lack of fusion. Similarly, Kwon et al. [37] enhanced YOLO by incorporating contextual information and a scale-aware mechanism, resulting in the CSA-YOLO model. This model outperformed the traditional YOLO across multiple evaluation metrics, such as precision, recall, and mean average precision (mAP), significantly improving the accuracy of welding defect detection.

To address challenges related to low contrast and significant variations in defect shapes and sizes, Pan et al. [38] designed a new backbone network, WD-YOLO, specifically for weld defect detection. By replacing the YOLO backbone with the NeXt backbone and adding a dual-attention mechanism, the model enhanced detection performance for targets of varying sizes. Additionally, Zhang et al. [39] proposed the S-YOLO model, which introduces full-dimensional dynamic convolution and a NAM attention mechanism to enhance feature representation in regions of interest. This approach effectively addresses challenges such as small target detection, occlusion, and overlap, achieving an 8.9% improvement in mAP compared to the original model.

2.2. Wavelet Transforms in Deep Learning

Wavelet transforms are a powerful tool for signal processing, known for their ability to decompose signals into components at various frequency scales. Unlike Fourier transforms, which only capture frequency information, wavelet transforms provide time-frequency localization, making them particularly effective in analyzing non-stationary signals and complex spatial patterns. This property has been exploited in various computer vision tasks, including image denoising, compression, and texture analysis [40].

In the context of radiographic inspection, researchers have adapted wavelet theory to address challenges in X-ray and γ-ray image analysis. For instance, Liu et al. [41] proposed the Dynamic Weights-Based Wavelet Attention Neural Network (DWWA-Net), which integrated dynamic wavelet transform convolutions (WTConv) and multiview attention to enhance weak defect detection and denoise strong background interference in industrial images. This approach achieved state-of-the-art precision improvements, with a 6.0% increase on the GC10-DET dataset and a 4.3% increase on the NEU dataset .

Recent advances have focused on wavelet-based network defect detection for industrial applications. For example, Guo et al. [42] proposed a novel CNN-wavelet fusion method that converts laser ultrasonic signals into scalograms for automated subsurface defect width quantification. This method achieved 98.5% validation accuracy and 100% experimental accuracy, demonstrating robust defect characterization without manual feature extraction. However, the small experimental dataset (four signals) and heavy reliance on simulated data raise concerns about the generalizability of this approach to real-world industrial scenarios with complex defect geometries and material variations.

Another notable contribution is the hybrid method proposed by Yang et al. [43], which integrated wavelet decomposition, Canny edge detection, and a multi-channel CNN to detect surface defects on automobile pipe joints. This method effectively addressed challenges posed by smooth surfaces and processing textures while achieving high classification accuracy under uneven illumination and noise. However, the dependency on manual parameter tuning in pre-processing stages may limit the scalability and applicability of this method in real-world industrial settings.

Despite significant advancements, conventional YOLO architectures rely on standard convolutional operations, which are limited in their ability to capture multi-scale and multi-frequency features inherent in radiographic images. This limitation becomes particularly evident when detecting defects with diverse morphologies, subtle grayscale variations, and complex edge boundaries. While most existing approaches treat wavelet transforms as auxiliary modules or preprocessing steps, achieving notable results, our work integrates wavelet transforms directly into the YOLO framework. This integration enables end-to-end multi-resolution feature extraction, specifically tailored for radiographic defect detection, allowing the network to dynamically extract multi-scale features and enhance detection accuracy. By doing so, our approach offers a promising direction for advancing automated defect recognition in industrial applications.

3. Materials and Methods

This section provides a detailed overview of the dataset utilized in this study and the preprocessing techniques employed to enhance its diversity. Additionally, the architecture of the proposed Wavelet Transform Yolo model (WT-YOLO) for welding flaw detection is described in detail, along with crucial training parameters and the experimental setup.

3.1. The Weld Defect Dataset and Pre-Processing

In this study, we utilized a dataset comprising 7000 radiographic images of weld defects. Each image is annotated with defect locations, though specific defect class labels are not provided. The dataset was divided into three subsets: a training set containing 4900 images, a validation set with 700 images, and a test set of 1400 images, which was used to evaluate the model's overall performance in detecting all types of defects.

Given the elongated nature of the weld images, a pre-processing strategy was implemented to segment the longer side of each image into smaller patches. Specifically, overlapping patches were extracted with a step size of 200 pixels, each with a fixed length of 640 pixels. These patches were then used for training, validation, and testing, ensur-

ing that the model learns from different regions of the image. This approach increases input diversity and enhances the model's robustness to variations in defect location and image geometry.

In addition to the general test set, we also created another test set consisting of 1400 defect images collected from on-site production environments, with each image containing only one type of defect. This set is designed to evaluate the model's ability to detect different defect categories. It includes seven defect classes: crack, lack of fusion, incomplete penetration, concavity, undercut, slag inclusion, and porosity, with 200 images per category. In order to facilitate a detailed performance comparison, defect images from each category are randomly sampled at a 10:1 ratio for manual assessment. This allows the model's predictions to be directly evaluated based on human detection.

To maintain consistency, all images, including the cropped patches, were resized to 640 × 640 pixels. Data augmentation techniques were applied to improve generalization, including random rotations (up to 30°), horizontal flipping, and random scaling. Furthermore, image normalization was performed by subtracting the mean and dividing by the standard deviation of pixel values across the dataset. These pre-processing steps collectively enhance the model's ability to detect defects across various scales and orientations.

3.2. Network Architecture

Given that YOLOv11 is accurate, efficient, and widely recognized, we selected it as the object detection model. The network architecture used in this study is based on the YOLOv11n framework [44], with a modification that replaces some of the traditional convolutional layers with wavelet transform convolutions (WTConv) [45] in the backbone. We refer to this modified model as WT-YOLO. This change is designed to enhance feature extraction capabilities, especially for the complex and elongated weld defects in the dataset. The architecture of the proposed method is illustrated in Figure 2.

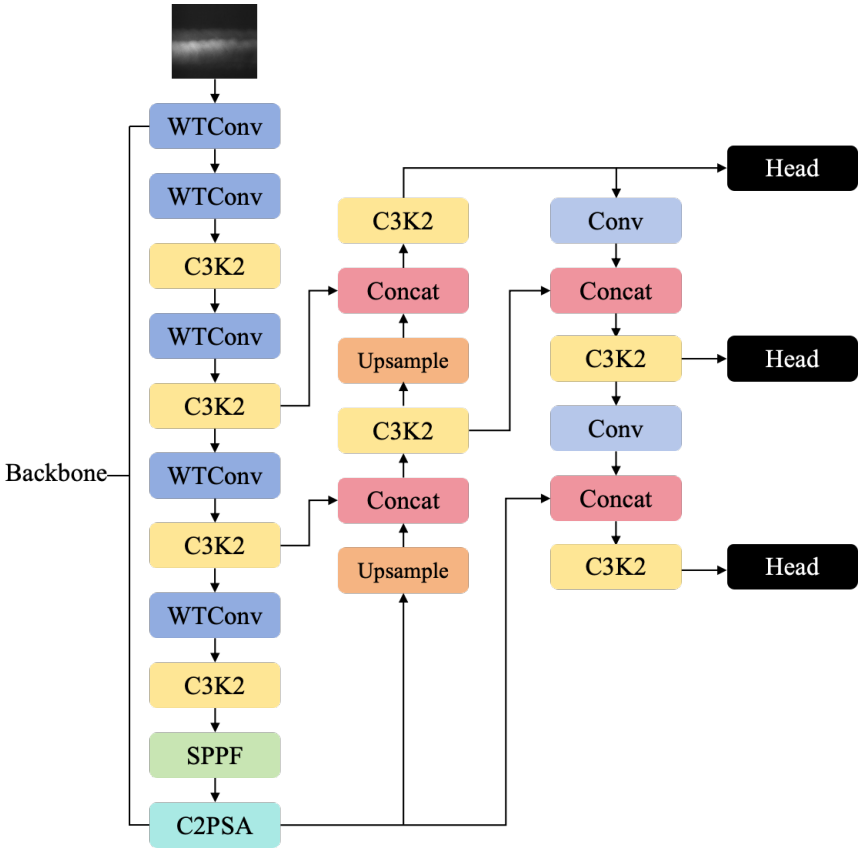


Figure 2. Structure of WT-YOLO for welding defect detection.

In the original YOLOv11n architecture, the backbone consists of a series of convolutional layers and residual blocks, followed by a detection head that predicts bounding boxes, class probabilities, and objectness scores. The detection head is responsible for classifying defects and localizing them within the image. During training, the network uses a combination of mean squared error for bounding box regression and cross-entropy loss for classification.

WT-YOLO's backbone is designed to extract hierarchical features from the input images and is based on a modified YOLOv11n architecture. The backbone begins with a standard convolution layer followed by a WTConv layer, which helps capture low-level features while integrating wavelet-based transformations for enhanced multi-scale feature extraction. As the network progresses, subsequent stages alternate between regular convolution and WTConv layers, allowing the model to learn progressively more complex and multi-scale representations of weld defects. The backbone's final stage incorporates a Spatial Pyramid Pooling (SPPF) layer to aggregate features at various scales, followed by a C2PSA layer that refines these features before passing them on to the head for further processing. This hierarchical processing enables the network to gradually reduce spatial resolution while increasing the number of feature channels, which in turn helps the model learn rich and diverse representations of the weld defects.

The head of the model leverages the features extracted by the backbone for prediction tasks. It begins with an upsampling operation that doubles the spatial resolution of the feature maps, followed by concatenation with the corresponding features from the backbone. This concatenation allows the network to retain high-resolution details from earlier layers. The upsampled features are then refined through a series of convolutional layers, processing them at different scales. The head concludes with a detection layer that outputs predicted bounding boxes and objectness scores. Overall, WT-YOLO architecture, with its combination of WTConv in the backbone and a multi-scale detection head, is optimized to detect fine-grained defect features efficiently, which is crucial for accurate weld defect detection across varying defect types and sizes.

3.3. WTConv Layer

WTConv is a key innovation in this study, designed to address the need for extracting multi-frequency features, which is crucial in the task of weld defect detection. Traditional convolutional layers in neural networks focus on extracting spatial patterns at a single scale, which may not effectively capture the wide range of frequency components present in welding defects. On the other hand, WTConv enable the network to simultaneously process different frequency components of the input image, both high- and low-frequency, by leveraging the wavelet transform. The structure of the WTConv layer is illustrated in Figure 3.

The wavelet transform is a mathematical tool that decomposes an image into different frequency bands. In contrast to Fourier transforms, wavelets have the ability to localize both in space and frequency, making them particularly effective at capturing localized features at different scales. The 2D continuous wavelet transform of an image $I(x, y)$ is given by

$$W_{a,b} = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} I(x, y) \psi_{a,b}(x, y)\, dx\, dy \tag{1}$$

where $\psi_{a,b}(x, y)$ represents the wavelet function and a, b are scale and translation parameters, respectively. The wavelet function has compact support, allowing it to respond only to localized features, making it ideal for capturing details in the spatial domain. In the context of weld defect detection, these localized features vary across frequency bands. For example,

cracks are often sharp, high-frequency features, while defects like concave or porosity tend to appear as smoother, low-frequency structures.

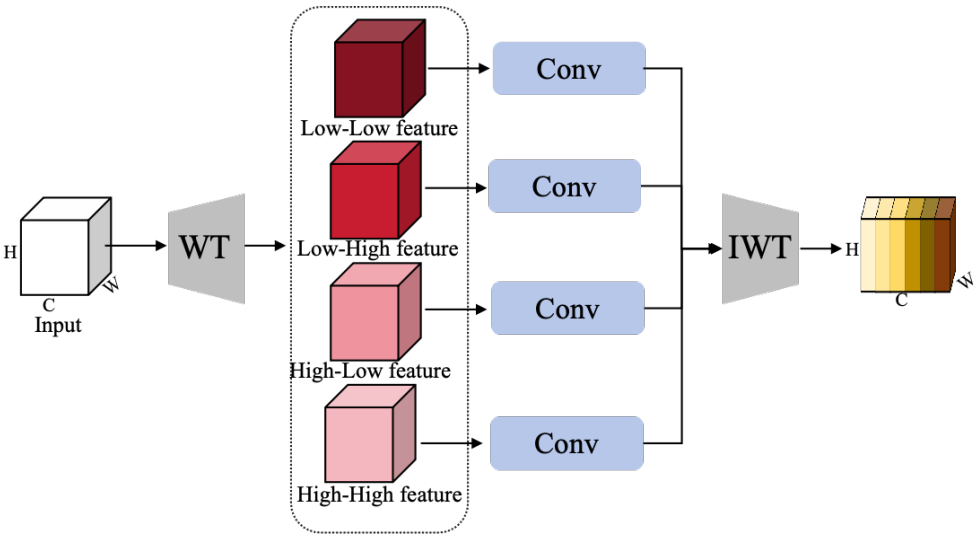


Figure 3. Structure of the WTConv layer. WT: Wavelet Transform. Low-Low: Contains low-frequency information in both horizontal and vertical directions, representing the overall structure and smooth regions of the image. Low-High: Contains low-frequency horizontal information and high-frequency vertical information, capturing vertical edges and textures in the image. High-Low: Contains high-frequency horizontal information and low-frequency vertical information, capturing horizontal edges and textures. High-High: Contains high-frequency information in both horizontal and vertical directions, capturing fine details and diagonal edges in the image. IWT: Inverse Wavelet Transform.

In the WT-YOLO model, the standard convolution operation is replaced with wavelet-based filters, which allows the network to perform localized frequency decomposition on the input feature maps. This operation is mathematically similar to regular convolution but uses wavelet functions as filters, enabling the network to capture features at multiple scales. The output of a WTConv operation can be expressed as

$$I_{out}(x,y) = \sum_{i,j} I(x+i, y+j) \cdot \psi_{a,b}(i,j) \tag{2}$$

where $\psi_{a,b}(i,j)$ is the wavelet filter applied to the input image $I(x,y)$, and the sum represents the convolution process. By using wavelets, WT-YOLO can efficiently extract both low-frequency global structures and high-frequency local details.

The application of WTConv in the context of weld defect detection provides several advantages. Different types of weld defects, such as cracks, slag inclusion, or porosity, manifest at different frequencies in the image domain. WTConv enables the network to detect these defects by processing both high- and low-frequency components simultaneously. Additionally, wavelets' ability to localize features in both spatial and frequency domains improves WT-YOLO's ability to identify defects at various scales, from sharp edges to smooth regions. Figure 4 shows the application of WTConv on cracks. This multi-frequency feature extraction not only enhances the model's robustness in detecting a wide range of defects but also improves its generalization ability, allowing it to perform well across different types of defects and welding conditions.

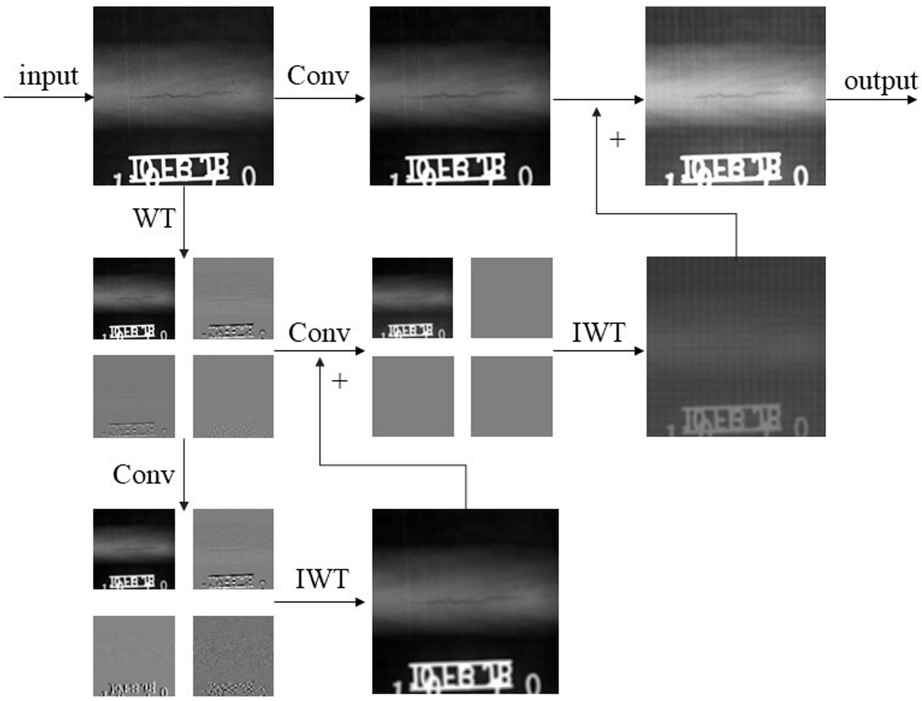


Figure 4. An example of WTConv operation on cracks.

YOLOv11n

In this study, the traditional convolutional layers of YOLOv11n are replaced with WTConv layers at critical points in the network architecture. These layers are inserted in the early stages of the backbone to decompose feature maps into both high- and low-frequency components. As the network progresses, the integration of these multi-frequency features allows it to more effectively capture both fine-grained details (such as cracks) and large-scale structures (such as incomplete penetration). This approach significantly improves WT-YOLO's capability in weld defect detection, making it more suitable for practical applications in industrial settings.

3.4. Experiment Settings

To assess the effectiveness of WTConv, we conducted an ablation study by systematically integrating WTConv layers into the baseline model (YOLOv11n) and analyzing their impact on model performance. The backbone module in YOLOv11n plays a crucial role in extracting abstract features of defects, which is why we replaced each convolutional layer in the backbone with WTConv to enhance feature extraction. The evaluation was performed using both the general test set and the specialized defect-specific test set to determine the model's capability in detecting different defect types.

All experiments were conducted using the PyTorch 2.5.1 framework on a GeForce RTX 4090 GPU with 24 GB of memory. Each model was trained for 400 epochs with a batch size of 16, utilizing the SGD optimizer with a learning rate of 0.005 according to the official recommendations of YOLOv11. To ensure a fair comparison, image pre-processing techniques such as cropping, resizing to 512 $\times$ 512, and data augmentation (rotations, flips, and scaling) were consistently applied across all experiments. The best-performing model parameters, determined based on validation set accuracy, were selected for the final evaluation.

For the defect-specific test set, additional performance metrics were recorded, including precision, recall, and F1-score for each defect class. Moreover, a subset of predictions was randomly selected in proportion to compare the model's outputs with human in-

spection results, providing a qualitative assessment of the model's detection accuracy in field applications.

3.5. Evaluation Criteria

To rigorously evaluate the performance of WT-YOLO in weld defect detection, we adopt standard object detection metrics that provide a balanced assessment of detection accuracy and reliability.

Mean Average Precision (mAP) [44] is used to measure the model's detection capability across different localization thresholds. Specifically, we report:

- mAP50–95: The mean of average precision values computed at IoU thresholds ranging from 0.50 to 0.95 with a step size of 0.05. This metric reflects the model's ability to consistently detect defects at varying localization accuracies.
- mAP50: The average precision at an IoU threshold of 0.50, commonly used in object detection tasks as a baseline measure of performance.
- mAP75: The average precision at an IoU threshold of 0.75, representing a more stringent requirement for precise localization.

Recall quantifies the model's ability to detect actual defects, calculated as

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}} \tag{3}$$

where TP denotes true positives and FN denotes false negatives. A higher recall indicates that fewer defects are missed.

Precision measures the proportion of correctly identified defects among all detections, given by

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}} \tag{4}$$

where FP represents false positives. A model with high precision makes fewer incorrect predictions.

F1-score is the harmonic mean of precision and recall, providing a balanced evaluation when considering both false positives and false negatives:

$$F1 = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \tag{5}$$

This metric is particularly useful when precision and recall need to be considered jointly, ensuring that the model is neither too conservative nor overly permissive in its detections.

By combining these metrics, we gain a comprehensive understanding of the model's performance, assessing not only its ability to detect defects but also its reliability in distinguishing true defects from false alarms.

4. Result

To comprehensively evaluate the impact of WTConv on the YOLOv11n model, we conducted extensive experiments on both the general test set and the specialized defect-specific test set. The analysis not only focuses on metric improvements but also delves into WT-YOLO's internal mechanisms, particularly the effects of WTConv on feature extraction, defect detection, and localization.

4.1. Overall Performance

Table 1 presents the overall detection performance comparison between the baseline YOLOv11n and WT-YOLO. WT-YOLO achieved improvements across all IoU thresholds,

with mAP50-95 increasing from 0.2397 to 0.2515, mAP50 from 0.5519 to 0.5675, and mAP75 from 0.1624 to 0.1836, demonstrating enhanced localization accuracy. Notably, precision improved from 0.58 to 0.6279, while recall slightly decreased from 0.5453 to 0.5400. The integration of WTConv introduces multi-scale frequency domain analysis, enabling WT-YOLO to capture both global and local texture patterns through multi-scale feature extraction. This capability is particularly beneficial for detecting fine-grained defect features, as it suppresses noise and amplifies meaningful structural details. The increased precision indicates that WTConv enhances the model's ability to differentiate defects from the background, reducing false positives. However, the slight drop in recall suggests that some defects with subtle intensity variations may not be fully captured. This is because frequency decomposition can alter spatial information.

Table 1. Performance metrics for the test set.

Metric	YOLOv11n	WT-YOLO
mAP50-95	0.2397	0.2515
mAP50	0.5519	0.5675
mAP75	0.1624	0.1836
Recall	0.5453	0.5400
Precision	0.5800	0.6279
F1	0.5621	0.5806

For 1400 defect images, WT-YOLO completed the detection in 28 s, averaging approximately 0.02 s per image, while the baseline YOLOv11n took 36 s, averaging about 0.026 s per image. This demonstrates the potential of WT-YOLO as a viable approach for real-time defect detection in industrial applications.

4.2. Defect-Specific Analysis

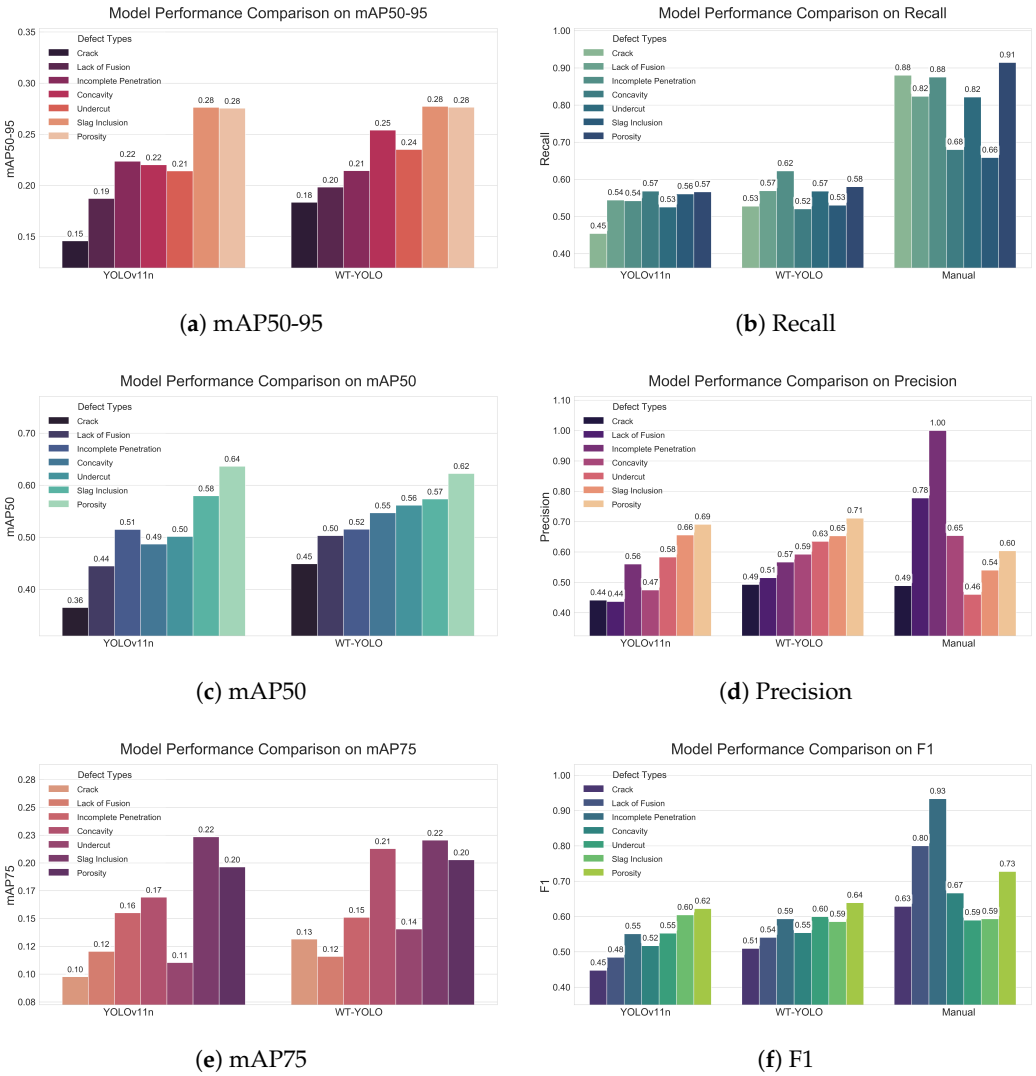
To further investigate the impact of WTConv on different defect types, we analyzed model performance across individual categories, as seen in Figure 5. The results highlight distinct patterns in how WTConv affects defect detection, localization, and false positive/negative rates for various defects.

Experiments were conducted on a dataset containing seven types of defects, with 200 images per defect category. The ablation study shows that WTConv leverages multi-scale frequency domain analysis to enhance the model's overall localization accuracy. In terms of detection accuracy, there is a slight decline for slag inclusion, while for other defect types—especially high-risk defects such as cracks, lack of fusion, and incomplete penetration—there are significant improvements.

In terms of localization accuracy, as shown in Figure 5a,c,e, the proposed WT-YOLO model shows significant improvements over YOLOv11n in terms of mAP50-95, mAP50, and mAP75 for defects such as cracks, concave, and undercut. For example, the three metrics for cracks were improved from 0.1458 to 0.1835, from 0.3649 to 0.4489, and from 0.0976 to 0.1314, respectively. For defects such as lack of fusion, incomplete penetration, slag inclusion, and porosity, the three metrics are generally similar, with the exception of porosity, where the mAP decreased by 0.0139, and other changes being less than 0.01.

In terms of detection accuracy for various defects, as shown in Figure 5b,d,f, such as cracks, lack of fusion, incomplete penetration, undercut, and porosity, the proposed model shows a comprehensive improvement in recall, precision, and F1-score compared to YOLOv11n. For example, the recall rate for cracks increased from 0.4539 to 0.5277, precision improved from 0.4411 to 0.4926, and the F1-score rose from 0.4474 to 0.5095. However, for slag inclusion, the recall rate decreased by 0.0304, precision decreased by 0.0028, and the

F1-score dropped by 0.0192. For concave defects, the recall rate decreased by 0.0480, while precision increased by 0.1180, and the F1-score increased by 0.0166.



(**a**) mAP50-95 (**b**) Recall

(**c**) mAP50 (**d**) Precision

(**e**) mAP75 (**f**) F1

Figure 5. Comparison of YOLOv11n, WT-YOLO and manual performance metrics for each class.

4.3. Comparison with Human Inspection

To systematically evaluate the practical performance of the proposed model, this study invited a radiographic testing expert with advanced certification from the State Administration for Market Regulation to independently conduct manual inspections on 140 radiographic films containing seven types of defects. The evaluation focused solely on defect detection performance, disregarding localization accuracy. The experiment compared the model and manual inspection in three key aspects: defect sensitivity, noise robustness, and efficiency.

Figure 5b,d,f shows that the proposed model significantly enhances sensitivity to low-contrast defects through frequency domain decomposition. In the detection of defects such as slag inclusion (Precision: 0.6529 vs. 0.5400), porosity (0.7112 vs. 0.6038), and undercut (0.6347 vs. 0.4600), WT-YOLO's precision exceeded that of manual inspection, with the largest difference reaching 0.1747 (for undercut). This indicates that WTConv's frequency-domain filtering effectively suppresses artifacts and background textures in radiographic films, reducing false positives. Additionally, for a total of 140 defect images, the model's inference time is only 10 s, averaging approximately 0.07 s per image, which is nearly 300 times faster than manual inspection, which takes over 80 min.

However, the model exhibits relatively lower sensitivity and recall rates for high-risk defects, such as cracks (Recall: 0.5277 vs. 0.8800), lack of fusion (0.5692 vs. 0.8235), and incomplete penetration (0.6225 vs. 0.8750). Additionally, the model's F1-score for complex edge defects, such as incomplete penetration (0.5933 vs. 0.9333), still indicates room for improvement. The comparative examples are shown in Figure 6.

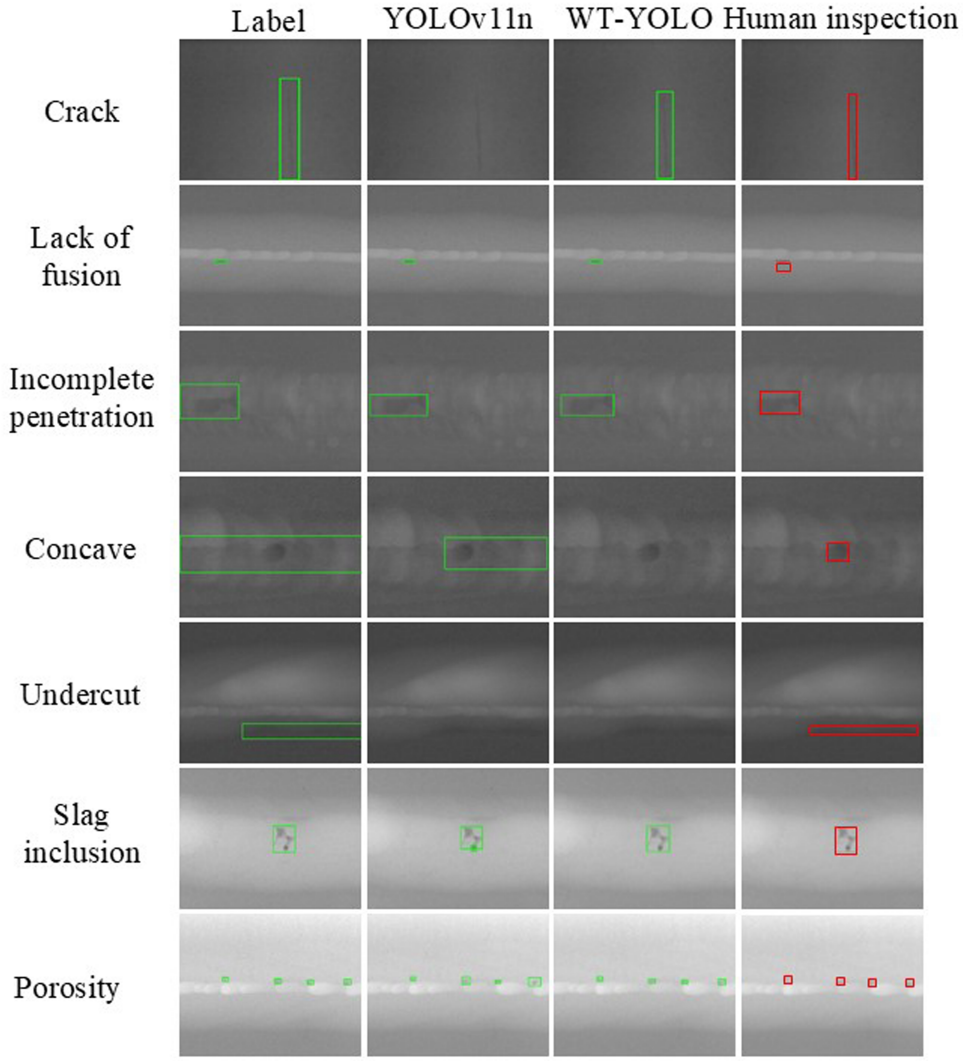


Figure 6. Typical defect detection result. The red and green boxes indicate the identified defect locations.

5. Discussion

The experimental results from defect-specific analysis demonstrate that WTConv significantly enhances the localization accuracy of WT-YOLO, particularly for high-risk defects such as cracks, lack of fusion, and incomplete penetration. The improvements in mAP50-95, mAP50, and mAP75 for these defects demonstrate that the multi-scale frequency domain analysis provided by WTConv effectively enables multi-scale feature extraction, which is critical for accurate defect localization. However, the slight decline in mAP for porosity indicates that WTConv may struggle with defects that exhibit subtle gray-scale variations, as frequency decomposition can sometimes alter spatial details. The improvements in recall, precision, and F1-score for cracks, lack of fusion, and incomplete penetration highlight the effectiveness of WTConv in extracting edge features and suppressing noise. The increased precision suggests that WTConv enhances the model's ability to differentiate defects from

background interference, reducing false positives. However, the slight decline in recall for slag inclusion and concave defects suggests that WTConv may miss some irregularly shaped defects or those relying on subtle shadow changes. This trade-off between precision and recall suggests that, while WTConv enhances overall detection accuracy, further refinement may be needed to fully capture all the defect types. The performance of WT-YOLO in detecting concave defects is particularly noteworthy. Although recall decreased, the improvements in precision and F1-score suggest that WTConv achieves a better balance between detecting true positives and minimizing false positives. This indicates that WTConv is particularly effective in scenarios where defect detection relies on subtle intensity variations, even if some defects are missed. The overall improvement in metrics across the seven defect categories demonstrates the robustness of WT-YOLO. Future work could focus on enhancing the model's ability to preserve spatial details during frequency decomposition, potentially through adaptive wavelet bases or hybrid frequency-space domain architectures.

The experimental results of comparison with humann inspection highlight the strengths and limitations of the WT-YOLO model in practical defect detection scenarios. The model's superior precision in detecting low-contrast defects, such as slag inclusion, porosity, and undercut, demonstrates the effectiveness of WTConv in suppressing noise and background interference. This is particularly valuable in industrial settings, where reducing false positives is critical for efficient batch screening. The model's inference speed, nearly 300 times faster than manual inspection, further highlights its potential for real-time applications, significantly improving operational efficiency. However, the lower recall rates for elongated defects, such as cracks, lack of fusion, and incomplete penetration, reveal a limitation of WTConv. While the frequency-domain filtering process effectively suppresses noise, it may result in the loss of spatial details. Compared to manual inspection, which leverages multi-scale visual focus and expert judgment, WT-YOLO struggles with detecting elongated defects, suggesting the need for additional mechanisms, such as spatial domain enhancement, to improve sensitivity. Future improvements could focus on integrating attention modules or similar spatial domain enhancement techniques to preserve critical spatial details. Notably, WT-YOLO excels in detecting blurred edge defects (e.g., undercut), while manual inspection remains superior for identifying elongated defects (e.g., lack of fusion). This indicates that a hybrid detection system combining the strengths of automated models and human expertise could offer a more comprehensive solution for industrial radiographic testing.

6. Conclusions

In this paper, compared to the baseline YOLOv11n, WT-YOLO is well-suited for detecting cracks, concavity, and undercut, as well as localizing defects such as cracks, incomplete penetration, lack of fusion, and porosity. When compared to manual inspection, WT-YOLO excels in the rapid and precise screening of undercut, slag inclusions, and porosity. Combining WT-YOLO with manual inspection could significantly enhance both the efficiency and accuracy of defect detection.

Author Contributions: Methodology, W.W. and Q.Z.;validation, J.P., L.Z. and H.C.; investigation, W.W. and H.C.;writing—original draft preparation, W.W.; writing—review and editing, W.W. and Q.Z.; visualization, W.W. and Q.Z. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported by Key Research and Development Program of Jiangxi Province (20224AAC01011).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The data presented in this study are available on request from the corresponding author. The data are not publicly available due to existing agreements with our collaborators.

Acknowledgments: The authors would like to thank the anonymous reviewers for the insightful discussions.

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

References

1. Messler, R.W., Jr. *Principles of Welding: Processes, Physics, Chemistry, and Metallurgy*; John Wiley & Sons: Hoboken, NJ, USA, 2008.
2. Phillips, D.H. *Welding Engineering: An Introduction*; John Wiley & Sons: Hoboken, NJ, USA, 2023.
3. Shravan, C.; Radhika, N.; NH, D.K.; SivaSailam, B. A review on welding techniques: Properties, characterisations and engineering applications. *Adv. Mater. Process. Technol.* **2023**, *10*, 80.
4. Buang, A.S.; Bakar, M.S.A.; Rohani, M.Z. A review of trend advanced welding process and welding technology in industries. *Int. J. Technol. Vocat. Eng. Technol.* **2024**, *5*, 133–145.
5. Zhao, L.; Tu, Y.; Gao, M. Research on the Effects of Welding Defects on the Strength Performance of Expandable Profile Liners and Approaches to Improve Welding Reliability. *SPE J.* **2023**, *28*, 540–553. [CrossRef]
6. Dwivedi, S.K.; Vishwakarma, M.; Soni, A. Advances and researches on non destructive testing: A review. *Mater. Today Proc.* **2018**, *5*, 3690–3698. [CrossRef]
7. Wang, B.; Zhong, S.; Lee, T.L.; Fancey, K.S.; Mi, J. Non-destructive testing and evaluation of composite materials/structures: A state-of-the-art review. *Adv. Mech. Eng.* **2020**, *12*, 1687814020913761. [CrossRef]
8. Sharma, K. Analysis of Non-destructive Testing for Improved Inspection and Maintenance Strategies. *e-J. Nondestruct. Test.* **2023**. [CrossRef]
9. Deepak, J.; Raja, V.B.; Srikanth, D.; Surendran, H.; Nickolas, M. Non-destructive testing (NDT) techniques for low carbon steel welded joints: A review and experimental study. *Mater. Today Proc.* **2021**, *44*, 3732–3737. [CrossRef]
10. Rossi Ciampolini, R.; Olsen, A.A. Ultrasonic Inspection Using Time-of-Flight Diffraction (TOFD). In *Testing and Inspection of Offshore and Marine Lifting Appliances: Class Requirements for Certification*; Springer: Berlin/Heidelberg, Germany, 2024; pp. 75–91.
11. Sweeney, N.E.; Parke, S.; Lines, D.; Loukas, C.; Vasilev, M.; Pierce, S.G.; MacLeod, C.N. In-process phased array ultrasonic weld pool monitoring. *NDT E Int.* **2023**, *137*, 102850. [CrossRef]
12. Otayni, A.; Aftan, A.A.; Nammazi, J.; Aljwear, A.; Mahnashy, A.; Al Dosary, M.; Hamlan, L.S.; Alkhthlan, M.; Mubaraki, A.; Saeed, M.K.; et al. Assessment of Rejected Radiographs during planar imaging procedures. *J. Radiat. Res. Appl. Sci.* **2023**, *16*, 100556. [CrossRef]
13. Wang, Y.; Zhang, Q. A characterization study on perovskite X-ray detector performance based on a digital radiography system. *Nucl. Sci. Tech.* **2023**, *34*, 69. [CrossRef]
14. Ghadarah, N.; Ayre, D. A review on acoustic emission testing for structural health monitoring of polymer-based composites. *Sensors* **2023**, *23*, 6945. [CrossRef]
15. Zang, X.; Xu, Z.D.; Lu, H.; Zhu, C.; Zhang, Z. Ultrasonic guided wave techniques and applications in pipeline defect detection: A review. *Int. J. Press. Vessel. Pip.* **2023**, *206*, 105033. [CrossRef]
16. Wang, D.; Zheng, Y.; Dai, W.; Tang, D.; Peng, Y. Deep network-assisted quality inspection of laser welding on power Battery. *Sensors* **2023**, *23*, 8894. [CrossRef] [PubMed]
17. Ramírez, D.P.; Veitía, B.D.R.; Ariosa, P.F.; Hernández, A.E.; Gilart, R.A.; Roca, Á.S.; Fals, H.D.C. Pore segmentation in industrial radiographic images using adaptive thresholding and Morphological analysis. *Trends Agric. Environ. Sci.* **2023**, *1*, e230008.
18. Prunella, M.; Scardigno, R.M.; Buongiorno, D.; Brunetti, A.; Longo, N.; Carli, R.; Dotoli, M.; Bevilacqua, V. Deep learning for automatic vision-based recognition of industrial surface defects: A survey. *IEEE Access* **2023**, *11*, 43370–43423. [CrossRef]
19. WANG, R.; HU, Y.; LIU, W.; LI, H. Defect detection of weld X-ray image based on edge AI. *Trans. China Weld. Inst.* **2022**, *43*, 79–84.
20. Kumaresan, S.; Aultrin, K.J.; Kumar, S.; Anand, M.D. Transfer learning with CNN for classification of weld defect. *IEEE Access* **2021**, *9*, 95097–95108. [CrossRef]
21. Amarnath, M.; Sudharshan, N.; Srinivas, P. Automatic detection of defects in welding using deep learning—A systematic review. *Mater. Today Proc.* 2023, *in press*.
22. Taye, M.M. Theoretical understanding of convolutional neural network: Concepts, architectures, applications, future directions. *Computation* **2023**, *11*, 52. [CrossRef]
23. Das, S.; Tariq, A.; Santos, T.; Kantareddy, S.S.; Banerjee, I. Recurrent neural networks (RNNs): Architectures, training tricks, and introduction to influential research. In *Machine Learning for Brain Disorders*; Humana: New York, NY, USA, 2023; pp. 117–138.

24. Yu, X.; Wang, S.; Hu, J. Guided Random Mask: Adaptively Regularizing Deep Neural Networks for Medical Image Analysis by Potential Lesions. *Appl. Sci.* **2022**, *12*, 9099. [CrossRef]
25. Patil, R.V.; Reddy, Y.P. Multiform weld joint flaws detection and classification by sagacious artificial neural network technique. *Int. J. Adv. Manuf. Technol.* **2023**, *125*, 913–943. [CrossRef]
26. Liu, T.; Zheng, P.; Bao, J. Deep learning-based welding image recognition: A comprehensive review. *J. Manuf. Syst.* **2023**, *68*, 601–625. [CrossRef]
27. Wang, R.; Wang, H.; He, Z.; Zhu, J.; Zuo, H. WeldNet: A lightweight deep learning model for welding defect recognition. *Weld. World* **2024**, *68*, 2963–2974. [CrossRef]
28. Zhao, Z.Q.; Zheng, P.; Xu, S.t.; Wu, X. Object detection with deep learning: A review. *IEEE Trans. Neural Netw. Learn. Syst.* **2019**, *30*, 3212–3232. [CrossRef] [PubMed]
29. 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, Las Vegas, NV, USA, 27–30 June 2016; pp. 779–788.
30. Wang, W.; Chen, J.; Han, G.; Shi, X.; Qian, G. Application of Object Detection Algorithms in Non-Destructive Testing of Pressure Equipment: A Review. *Sensors* **2024**, *24*, 5944. [CrossRef]
31. Yu, Y.; She, K.; Liu, J.; Cai, X.; Shi, K.; Kwon, O.M. A super-resolution network for medical imaging via transformation analysis of wavelet multi-resolution. *Neural Netw.* **2023**, *166*, 162–173. [CrossRef]
32. Wang, C.Y.; Liao, H.Y.M. YOLOv1 to YOLOv10: The fastest and most accurate real-time object detection systems. *Apsipa Trans. Signal Inf. Process.* **2024**, *13*, e29. [CrossRef]
33. Khanam, R.; Hussain, M. Yolov11: An overview of the key architectural enhancements. *arXiv* **2024**, arXiv:2410.17725.
34. Redmon, J.; Farhadi, A. YOLO9000: Better, faster, stronger. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, HI, USA, 21–26 July 2017; pp. 7263–7271.
35. Redmon, J.; Farhadi, A. Yolov3: An incremental improvement. *arXiv* **2018**, arXiv:1804.02767.
36. Cheng, S.; Yang, H.G.; Xu, X.Q. Improved Lightweight X-Ray Aluminum Alloy Weld Defects Detection Algorithm Based on YOLOv5. *Chin. J. Lasers* **2022**, *49*, 9.
37. Kwon, J.E.; Park, J.H.; Kim, J.H.; Lee, Y.H.; Cho, S.I. Context and scale-aware YOLO for welding defect detection. *NDT E Int.* **2023**, *139*, 102919. [CrossRef]
38. Pan, K.; Hu, H.; Gu, P. Wd-yolo: A more accurate yolo for defect detection in weld X-ray images. *Sensors* **2023**, *23*, 8677. [CrossRef]
39. Zhang, Y.; Ni, Q. A novel weld-seam defect detection algorithm based on the s-yolo model. *Axioms* **2023**, *12*, 697. [CrossRef]
40. Guo, T.; Zhang, T.; Lim, E.; Lopez-Benitez, M.; Ma, F.; Yu, L. A review of wavelet analysis and its applications: Challenges and opportunities. *IEEE Access* **2022**, *10*, 58869–58903. [CrossRef]
41. Liu, J.; Zhao, H.; Chen, Z.; Wang, Q.; Shen, X.; Zhang, H. A dynamic weights-based wavelet attention neural network for defect detection. *IEEE Trans. Neural Netw. Learn. Syst.* **2023**, *35*, 16211–16221. [CrossRef] [PubMed]
42. Guo, S.; Feng, H.; Feng, W.; Lv, G.; Chen, D.; Liu, Y.; Wu, X. Automatic quantification of subsurface defects by analyzing laser ultrasonic signals using convolutional neural networks and wavelet transform. *IEEE Trans. Ultrason. Ferroelectr. Freq. Control* **2021**, *68*, 3216–3225. [CrossRef]
43. Yang, Z.; Zhang, M.; Li, C.; Meng, Z.; Li, Y.; Chen, Y.; Liu, L. Image classification for automobile pipe joints surface defect detection Using wavelet decomposition and convolutional neural network. *IEEE Access* **2022**, *10*, 77191–77204. [CrossRef]
44. Rasheed, A.F.; Zarkoosh, M. YOLOv11 Optimization for Efficient Resource Utilization. *arXiv* **2024**, arXiv:2412.14790.
45. Finder, S.E.; Amoyal, R.; Treister, E.; Freifeld, O. Wavelet convolutions for large receptive fields. In Proceedings of the European Conference on Computer Vision, Online Event, 29 September–4 October 2024; Springer: Berlin/Heidelberg, Germany, 2024; pp. 363–380.

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.