POLITECNICO DI TORINO
Repository ISTITUZIONALE

BW-YOLO: A Lightweight Recurrence-Plot-Driven Network for In-Process Defect Identification from Multi-Band Photodiode Signals in Lithium-Ion Battery Laser Welding

Original
BW-YOLO: A Lightweight Recurrence-Plot-Driven Network for In-Process Defect Identification from Multi-Band Photodiode Signals in Lithium-Ion Battery Laser Welding / Bao, N., Li, J., Fan, Y., Simeone, A., Bao, R.. - In: JOURNAL OF MANUFACTURING AND MATERIALS PROCESSING. - ISSN 2504-4494. - 10:11(2026). [10.3390/jmmp10110459]

Availability:
This version is available at: 11583/3017891 since: 2026-11-13T11:07:42Z

Publisher:
MDPI

Published
DOI:10.3390/jmmp10110459

Terms of use:

This article is made available under terms and conditions as specified in the corresponding bibliographic description in the repository

Publisher copyright

(Article begins on next page)

21 November 2026

Journal of
Manufacturing and Materials Processing

MDPI

Article

BW-YOLO: A Lightweight Recurrence-Plot-Driven Network for In-Process Defect Identification from Multi-Band Photodiode Signals in Lithium-Ion Battery Laser Welding

Nengsheng Bao [1], Jiaqi Li [2], Yuchen Fan [3], Alessandro Simeone [3,*] and Runxuan Bao [4]

[1] Intelligent Manufacturing Key Laboratory of Ministry of Education, College of Engineering, Shantou University, Shantou 515063, China; nsbao@stu.edu.cn
[2] QKM Robotics Co., Ltd., Dongguan 523000, China; 19jqli@alumni.stu.edu.cn
[3] Department of Management and Production Engineering, Politecnico di Torino, Corso Duca degli Abruzzi 24, 10129 Torino, Italy; yuchen.fan@polito.it
[4] College of Engineering, Guangdong Technion-Israel Institute of Technology, Shantou 515063, China; bao12180@gtiit.edu.cn
* Correspondence: alessandro.simeone@polito.it

Abstract

Laser welding of tabs and busbars determines the electrical integrity and safety of lithium-ion battery modules, yet many weld defects leave little trace on the surface that post-weld vision inspection relies on. Photodiode signals record the weld-pool dynamics during processing, but their short, non-stationary waveforms are difficult to classify reliably with a compact model. This study presents BW-YOLO, a lightweight network designed together with recurrence-plot (RP) encoding of the photoelectric signal. The RP turns each weld signal into an image whose coarse block structure and fine recurrent texture reflect keyhole stability. BW-YOLO matches this two-scale structure with a Haar wavelet convolution module that integrates channel shuffle and one-shot aggregation (WT-ShuffleOSA) and a dual-scale channel reduction pyramid slice attention module (DCR-PSA). The evaluation used a balanced dataset of 1200 industrially produced joints monitored in the visible, infrared and reflected-laser bands, with one fixed training–validation partition. On the 240 validation joints, evaluated separately in each band, BW-YOLO reduced misclassifications by 30–38% in every band relative to the YOLO11n-cls baseline with a reported 15% reduction in operation count. In the infrared band it reached 96.67% accuracy and 98.28% precision with 1.50 M parameters and 2.8 GFLOPs, giving the highest observed accuracy among the ten networks compared on that partition. These results support further investigation of the proposed design; robustness across training runs and datasets, performance relative to direct time-series classifiers, and deployment latency remain to be established.

Keywords: laser welding; lithium-ion battery; in-process monitoring; photodiode signal; recurrence plot; wavelet convolution; lightweight neural network; defect identification

Check for updates

Academic Editor: Yu-Ming Zhang

Received: 14 October 2026
Revised: 22 October 2026
Accepted: 23 October 2026
Published: 30 October 2026

Copyright: © 2026 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.

1. Introduction

Worldwide demand for lithium-ion batteries has grown rapidly with the electrification of transportation and the expansion of grid-scale energy storage. Global battery manufacturing capacity exceeded 4 TWh by the end of 2025, an increase of roughly 30% over the previous year, with year-on-year capacity growth exceeding 50% in the European Union and the United States [1]. Lithium-ion batteries combine high energy density, long cycle life and low self-discharge, properties that have made them the dominant technology across

electric vehicles, portable electronics and grid-scale energy storage [2]. Meeting this scale of demand while maintaining consistent product quality places increasing pressure on battery manufacturing processes, where even a small increase in the defect rate translates into a large absolute number of rejected or unsafe cells at gigafactory production volumes.

Within a battery module, the current-carrying path depends on laser welding of electrode tabs to current collectors and of cell tabs to busbars, a joining step that must accommodate dissimilar metals such as copper, aluminum and nickel while keeping contact resistance low and mechanical strength high [3,4]. Laser welding is preferred for this task because of its deep penetration, high precision, small heat-affected zone and flexible processing, and it has become a dominant joining technology across electric vehicle battery production lines [5]. Weld quality nonetheless remains sensitive to parameter settings, surface contamination, material variability and environmental conditions, and the resulting defects are not always confined to the weld surface: porosity, incomplete fusion and unstable keyhole behavior can originate below the surface while leaving little visible trace. A defective joint can raise local contact resistance, generate excess heat during charging and discharging, and in severe cases initiate thermal runaway, which translates into yield loss and costly recalls [6].

Offline methods, including metallographic sectioning, scanning electron microscopy and computed tomography, remain the reference standard for characterizing weld morphology, but they are destructive or time-consuming and can only be applied to sampled joints. Two broad paradigms have therefore been used to control this risk. Statistical and gate-based quality management divides the production line into checkpoints at which measured process or product characteristics must satisfy predefined criteria before a cell is allowed to proceed to the next stage; this approach limits the propagation of defective components downstream but depends on the checkpoints being informative enough to catch process-level anomalies as they occur [7]. Sensor-based process monitoring instead aims to characterize every weld directly, using signals acquired during welding to make a pass or fail decision without interrupting the production cycle. Because photoelectric sensing of the weld pool is non-contact, low in cost and compatible with the cycle time of automated welding stations, it has become one of the most widely investigated sensing modalities for this purpose [8,9].

Turning photodiode signals into reliable inline decisions is shaped by three practical factors. First, the photoelectric signals generated during welding are strongly non-stationary: their amplitude, dominant frequency and variance drift within the short duration of a single weld as the keyhole forms, stabilizes and collapses, a dynamic behavior also reported for other real-time process signals such as arc and penetration-state measurements [10,11], so that a feature or threshold calibrated on one phase of the signal need not remain valid across the rest of it. Second, defects such as spatter-induced pitting and false or incompletely fused joints frequently leave only faint and easily confounded traces in the raw time-domain waveform, a difficulty also reported for infrequent or subtle failure modes in other welding processes [12,13], so that acceptable and defective welds can appear nearly indistinguishable unless the signal is re-expressed in a representation that amplifies their structural differences. Third, because the inspection module operates in line with the welding station, the recognition decision must be returned within the cycle time of the production line; recent reviews of real-time weld monitoring have identified this latency budget, rather than recognition accuracy alone, as the main barrier to moving learned models from offline benchmarks into deployed welding lines [14,15].

The broader adoption of deep learning in manufacturing has reshaped how these factors can be addressed. Learned feature representations have increasingly replaced handcrafted statistical features in process-monitoring tasks, and convolutional architectures

originally developed for natural image recognition have been adapted for signal classification, defect segmentation and quality prediction across a wide range of manufacturing processes [16]. For photoelectric weld monitoring, this shift raises two linked design questions: how a one-dimensional process signal should be represented so as to expose the features that separate acceptable from defective welds, and how a classification network should be structured to exploit that representation within an inline computational budget. This study answers both questions with a single design principle: the signal representation fixes the texture statistics that the classifier has to resolve, so the classifier should be built around those statistics. The resulting network, BW-YOLO, pairs recurrence-plot encoding of multi-band photodiode signals with wavelet-domain feature extraction and channel-reduced attention, and it is more accurate and less computationally expensive than its YOLO11 baseline on all three optical bands.

The remainder of this paper is organized as follows. Section 2 reviews related work and states the research gap and contributions. Section 3 describes the welding platform, the dataset, the recurrence-plot encoding and the BW-YOLO architecture. Section 4 reports the experimental results, Section 5 discusses their implications for inline battery-weld monitoring, and Section 6 concludes the paper.

2. Related Work

2.1. Photoelectric Sensing and Learning-Based Weld Monitoring

Among the sensing modalities investigated for online weld monitoring, including infrared thermography, acoustic emission and high-speed imaging of the melt pool, optical sensing of the light emitted from the weld pool has received particular attention because it is non-contact, structurally simple and compatible with the sampling rates required to characterize a millisecond-scale welding process [17]. Photodiode-based systems capture the visible-light, infrared and laser-reflection radiation emitted from the weld pool, and these signals correlate closely with penetration depth, spatter and keyhole stability [3,18]. This correlation has recently been confirmed for lithium-ion battery assembly, where photodiode signal analysis was used to monitor the laser beam welding of cell-internal contacts and to relate the recorded optical signatures to joint quality [8]. Commercial systems, such as the laser welding monitor developed by Precitec, compare the light emission of a new weld against a template built from qualified samples and issue a pass or fail decision using a fixed threshold [19], and a similar template and regression-based approach has been used to estimate weld quality in real time for automotive steel welding from statistical features of the photodiode signal [20]. This template-and-threshold strategy performs adequately for a fixed material and process window, but the template must be rebuilt whenever the material batch or welding parameters change, which is difficult to sustain in the small-batch, high-mix production that is common in battery manufacturing, and the hand-selected signal statistics may not capture the full information content of the sensor signal.

Learning-based approaches reduce this dependence on fixed thresholds. For copper-to-steel battery tabs, supervised classifiers trained on statistical and wavelet features of photodiode signals have separated lack of connection, sound welds and over-penetration [9], and wavelet scattering combined with a one-dimensional convolutional autoencoder has been used to flag anomalous welds from the same type of signal without labels [21]. Unlabeled photodiode data collected on an industrial welding platform has also been combined with labeled laboratory data within a deep learning framework to improve quality prediction from multi-channel photoelectric signals, targeting the laboratory-to-industry generalization gap [22]. Beyond photodiode monitoring, a deep belief network has been used to classify keyhole penetration state directly from process signals during laser welding [10]; for ultrasonic metal welding, sensor-fusion frameworks [23] and classical machine learning models

built on process-informed time-series features [24] outperform single-threshold decision rules; and deep networks with cross-attention have automated ultrasonic monitoring of resistance spot welds [25]. A review of real-time sensing in gas metal arc welding similarly concludes that dynamic, evolving process signals are poorly served by static thresholds calibrated on a fixed operating point [11]. These findings across welding processes motivate representations that are learned rather than manually specified.

2.2. Time-Series-to-Image Encoding for Signal-Based Defect Identification

One-dimensional convolutional networks and recurrent architectures can learn discriminative features directly from raw time series and have been used to fuse multiple photodiode channels through cross-attention for penetration state monitoring [26], to interpret photodiode signals recorded during ultrashort-pulsed laser structuring [27], and to couple multi-sensor diagnosis with machine learning for in situ monitoring of laser hot-wire cladding quality [28]. A wavelet convolutional network has likewise been applied to non-stationary acoustic emission signals for online identification of interfacial bonding quality during laser directed energy deposition, supporting the applicability of wavelet-domain feature extraction to non-stationary manufacturing process signals [29]. These studies demonstrate that direct time-series modelling is a viable approach to process monitoring. Its relative performance compared with image-based representations depends on the signal characteristics, architecture and training procedure and requires comparison under matched evaluation conditions.

An alternative line of work re-represents a time series as a two-dimensional image so that convolutional architectures developed for computer vision, together with associated pretraining and transfer learning practices, can be applied directly. The recurrence plot, first proposed to visualize the recurrence behavior of dynamical systems in phase space [30], measures the pairwise distance between reconstructed phase-space states of a signal and renders the result as a two-dimensional matrix, thereby exposing periodicity, drift and chaotic structure that are not easily observed in the original one-dimensional form; deep convolutional networks trained on such recurrence-plot images have been shown to classify time series accurately across a range of application domains [31]. Related encodings, including the Gramian angular field and the Markov transition field [32], have improved the classification of satellite image time series [33] and recognized weld defects from laser-scanned seam trajectories with higher accuracy than direct classification of structured-light images [34]. In processes other than welding, image-domain classification of infrared thermal fields has predicted injection-molding quality [35], and image-based detectors have recognized subtle surface defects in machined cylinder bores [36]. Across these studies, no single encoding is uniformly reported as superior, which suggests that the most effective representation depends on the statistical character of the specific signal being encoded and should be verified empirically.

2.3. Lightweight Networks for Weld Inspection

Once a process signal has been converted into an image, the choice of classification architecture determines both recognition accuracy and inference cost. The You Only Look Once (YOLO) family is widely used in manufacturing because it combines competitive accuracy with fast inference, and successive YOLO releases have been adapted for a range of surface and process inspection tasks [37]. An improved YOLOv5 architecture has been used to recognize defect signatures in weld radiographs [13], and a stable-diffusion-augmented YOLOv5 variant has been proposed for defect detection on stamped metal parts under limited training data [38]. Weld surface defects have been segmented online with a lightweight VGG16-UNet coupled to laser scanning [39], and lightweight convolutional networks have

been proposed for inspecting laser-welded safety vents on power battery cases [6], for real-time defect inspection with an edge-deployable model on power battery laser welding lines [40], and for automated quality inspection of laser-welded battery joints from vision data [41]. Attention mechanisms are increasingly used to weight the most informative parts of welding sensor data [42].

These studies operate on surface, profile or radiographic images and are effective at identifying visible geometric defects, but sub-surface weld anomalies that leave little trace on the surface are better reflected in the optical radiation emitted during welding than in a post-weld image. Applying an image-oriented network such as YOLO to recurrence-plot-encoded photoelectric signals calls for architectural adaptation beyond retraining on a new image type: the detection neck used for localizing objects in natural images is unnecessary for whole-image classification, and the fixed-scale convolutional kernels and standard self-attention layers of general-purpose backbones are not designed for the quasi-periodic texture that a recurrence plot produces or for the parameter and latency budgets of inline industrial deployment.

2.4. Research Gap and Contributions

Table 1 summarizes representative studies according to the signal or data type used, the representation applied before classification, the model type, and the main limitation relative to the requirements of battery laser welding inspection identified in Section 1.

Table 1. Representative studies on sensor-based and image-based weld quality inspection.

Study	Signal/Data Type	Representation	Model Type	Key Limitation Relative to This Study
Precitec monitor [19]	Photodiode signal	Raw amplitude vs. threshold	Template matching	Threshold must be rebuilt for each material or process change
Park et al. [20]	Photodiode signal	Statistical features	Empirical regression	Hand-selected features; validated for CO_2 laser welding of steel
Chianese et al. [9]	Photodiode signal (battery tabs)	Statistical and wavelet features	Supervised ML classifiers	Hand-crafted features; no learned 2D representation
Cai et al. [10]	Process signal	Raw signal features	Deep belief network	Not designed for image encoding or lightweight inline deployment
Cao et al. [26]	Multi-channel photodiode	Raw 1D signal	1D-CNN with cross-attention	Direct time-series approach; RP-based classification not evaluated
Geiger et al. [27]	Photodiode signal	Raw 1D signal	Statistical/ML classifier	Applied to laser structuring rather than battery welding
Chen et al. [28]	Multi-sensor signals	Raw 1D signal	Multi-sensing ML fusion	Applied to laser hot-wire cladding rather than battery welding
Hatami et al. [31]	General time series	Recurrence plot	Deep CNN	Not applied to welding signals or adapted for lightweight design
Wang and Oates [32]	General time series	GAF/MTF	CNN	Encoding not validated for photoelectric weld signals
Liu et al. [34]	Weld seam trajectory	2D image encodings	Deep neural network	Applied to seam geometry, not process-emitted signals

Table 1. *Cont.*

Study	Signal/Data Type	Representation	Model Type	Key Limitation Relative to This Study
Xu et al. [13]	Radiographic image	Direct image input	Improved YOLOv5	Operates on radiographs rather than an encoded process signal
Sun et al. [39]	Weld surface image	2D image	VGG16-UNet	Surface-level; cannot capture sub-surface anomalies
Yang et al. [6]	Weld surface image (battery vent)	2D image	Lightweight CNN	Vision-based; backbone not adapted for signal texture
This study (BW-YOLO)	Multi-band photodiode signal	Recurrence plot (compared with three alternatives on the infrared band)	Lightweight YOLO11-based (WT-ShuffleOSA + DCR-PSA)	Encoding and architecture designed together for in-process battery-weld monitoring

As Table 1 shows, existing studies address the representation problem and the efficiency problem largely in isolation. Threshold and regression-based photodiode methods operate directly on statistical features of the raw signal and avoid the cost of an image-based backbone, but they depend on hand-selected features and material-specific calibration [9,19,20]. Direct time-series and multi-sensor learning approaches provide alternatives to image-based classification for interpreting process signals [26–28]. Time-series-to-image encodings expose local structure by mapping the signal into two-dimensional space, but the studies that introduce or compare these encodings do so for general time series or non-welding applications, and it remains unverified which encoding best suits the short, non-stationary photodiode signals produced during battery laser welding [31–34]. Lightweight image-based networks, in turn, have been developed for post-weld surface images rather than for encoded process signals, and their backbones are not adapted to the texture statistics that a signal-to-image encoding produces [6,39,41]. A recent cross-process review of real-time weld-penetration monitoring likewise identifies the joint achievement of representation adequacy and inference efficiency, rather than either alone, as the main open challenge for deploying learned monitoring models [14].

This study addresses that gap with BW-YOLO (Battery Welding YOLO), a lightweight classification network for identifying acceptable and defective welds in lithium-ion battery laser welding from photoelectric signals. The contributions are as follows:

1. A representation–architecture co-design for photodiode-based weld monitoring. Recurrence-plot images of weld signals combine a coarse block structure with fine recurrent texture. BW-YOLO is built around this two-scale structure: Haar wavelet convolution separates the two scales explicitly, and channel-reduced attention links distant time instants of the weld at low cost.
2. Two modules tailored to the encoded signal texture. WT-ShuffleOSA builds on wavelet transform convolution, channel shuffle and one-shot aggregation. DCR-PSA builds on channel reduction attention and a dual-scale gated feed-forward unit. The novelty lies in how these operators are arranged and where they are placed. WT-ShuffleOSA applies wavelet convolution to half of the channels in each unit and mixes the two halves by channel shuffle. It also uses wavelet convolution for every convolution in the one-shot aggregation path. DCR-PSA keeps the split structure of the YOLO11 C2PSA block but replaces its self-attention and feed-forward network with these two units. The combined configuration achieved higher validation accuracy and a lower reported operation count than the YOLO11n-cls baseline in the experiments presented here.

3. An industrial multi-band evaluation. Using 1200 joints welded on a battery-module welding platform and monitored simultaneously in the visible, infrared and reflected-laser bands, the study compares RP with three alternative encodings on the infrared band, evaluates BW-YOLO against the baseline across all three bands, and benchmarks it against nine additional classification networks on a fixed validation partition. The reflected-laser band achieved the highest observed accuracy among the individual channels under the investigated conditions.

3. Materials and Methods

This section describes the welding platform and photoelectric sensing configuration, the dataset, the recurrence-plot encoding of the raw signals, the BW-YOLO architecture, and the training and evaluation protocol.

3.1. Welding Platform and Photoelectric Sensing

The signals were collected on an industrial battery-module laser welding platform (Figure 1). A quasi-continuous fiber laser (uw-S150Q-UI, United Winners Laser, Shenzhen, China) delivers the beam through a two-axis scanning galvanometer (SUPERSCAN IV-20 QU, RAYLASE, Wessling, Germany), which moves the spot along the programmed path on the cell tabs and busbar of the module. The galvanometer positions the beam by mirror deflection instead of mechanical axes, which gives fast, repeatable trajectories with low tracking error. The main laser specifications and the process parameters used in this study are listed in Table 2. All joints were welded under the same process conditions so that the recorded signals are directly comparable.

Table 2. Laser specifications and welding process parameters.

Item	Value
Laser wavelength	1070 nm
Maximum average power, continuous mode	250 W
Average power, pulsed mode	150 W
Maximum peak power, pulsed mode	1500 W
Maximum pulse energy	15 J
Pulse width, pulsed mode	0.2–50 ms
Power stability	$\leq \pm 1\%$
Operating mode used	Continuous
Laser power used	225 W
Welding speed	410 mm/s
Working distance	240 mm

The radiation emitted from the weld zone was acquired with a multi-channel laser-welding optical sensor (Diligine Optical Technology, Guangzhou, China) attached to the welding head. The sensor integrates three photodiodes behind optical filters that separate the visible band (400–700 nm), the infrared band (900–1700 nm) and the reflected-laser band (1070 ± 10 nm), so that the three signals are recorded synchronously for every joint. The acquisition unit applies low-pass filtering and opto-electrical isolation to suppress interference from laser modulation, power-supply coupling and the factory environment, and each channel is sampled at 20 kHz with 12-bit resolution.

The three bands carry physically different information. The visible signal is dominated by plasma-plume radiation and by light emitted or scattered by metal vapor and spatter, and its fluctuations reflect process stability. The infrared signal is dominated by thermal radiation from the melt pool and its surroundings, so its intensity follows the thermal behavior of the pool, and abnormal fluctuations are associated with false welds, porosity

or cracks. The reflected-laser signal, at the laser wavelength itself, records the part of the incident beam reflected by the melt-pool surface, the keyhole wall and spatter: it stays low and steady while the keyhole is open and stable, and it changes abruptly when the keyhole collapses or closes or when the pool oscillates strongly.

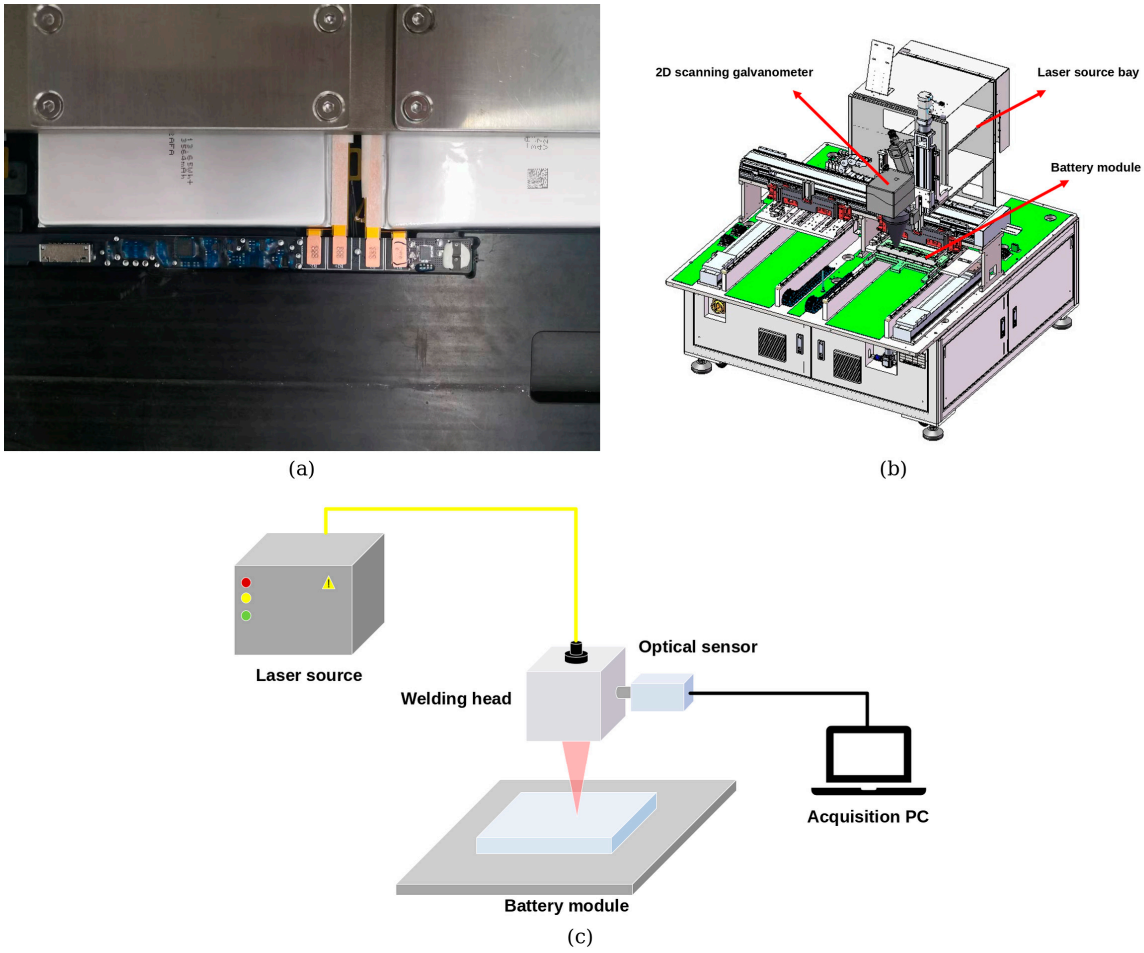


Figure 1. Experimental platform: (**a**) tab-to-busbar joints of the battery module; (**b**) battery-module laser welding platform; (**c**) schematic of the welding and photoelectric sensing configuration.

3.2. Dataset and Labeling

The study dataset comprised 1200 joints selected from production over approximately three months, including 600 acceptable weld joints and 600 defective weld joints.

For each joint, the visible, infrared and reflected-laser signals were recorded simultaneously, giving 3600 raw time series of approximately 400 sampling points each, which corresponds to about 20 ms per weld at the 20 kHz sampling rate. The quality of every joint was determined after welding by visual inspection followed by a destructive pull test. A joint without visible defects such as spatter, burn-through or weld misalignment and with a pull-off force of at least 30 N was labeled as an acceptable weld joint (AW); a joint with a visible defect or a pull-off force below 30 N was labeled as a defective weld joint (DW). The label therefore reflects the overall structural quality of the joint, which is the target of inline pass/fail inspection. The 30 N threshold is the acceptance criterion in the in-house quality standard of the industrial partner for this tab-to-busbar joint. All joints were welded with the nominal parameters in Table 2, and the defects occurred naturally during production. The dataset therefore contains only defects that arise in real production. The defective joints were picked out from this production output. The defects were not

divided into types, because the line only needs a pass/fail decision. The same number of acceptable joints from the same period was then selected to form a balanced dataset. This selection gave the two classes equal representation in training and validation. The resulting 50% defective-joint proportion reflects the dataset construction and should not be interpreted as the production-line defect rate.

Because the emission characteristics of the three bands are governed by different physical mechanisms, the three band-specific datasets were treated independently throughout data processing and network evaluation, which also allowed the contribution of each optical band to defect identification to be examined separately. Each single-band dataset is balanced, with 600 AW and 600 DW samples. For each band, samples were stratified by class and divided at the welded-joint level into training (80%) and validation (20%) subsets before the image representations were generated, so that data from the same joint never appear in both subsets. The same fixed partition was used for every experiment, so that all methods were evaluated on identical validation samples. Table 3 summarizes the composition of the datasets.

Table 3. Composition of the photoelectric signal datasets.

Optical Band	AW Samples	DW Samples	Total	Training (80%)	Validation (20%)
Each band (visible, infrared, reflected laser)	600	600	1200	960	240

3.3. Recurrence-Plot Encoding of Photoelectric Signals

A one-dimensional photodiode signal expresses the temporal evolution of the weld pool but represents its non-linear and non-stationary characteristics only implicitly. Recurrence-plot encoding reveals these characteristics by comparing the states of the signal at every pair of time instants and showing when the signal returns to a previously visited state [30]. Given a signal $X = \{x_1, \ldots, x_n\}$, the phase-space states are reconstructed by time-delay embedding:

$$\mathbf{s}_i = \left(x_i,\, x_{i+\tau},\, \ldots,\, x_{i+(m-1)\tau}\right), \quad i = 1, \ldots, N, \quad N = n - (m-1)\tau \tag{1}$$

where m is the embedding dimension and τ is the delay. The recurrence matrix compares every pair of states:

$$R_{i,j} = H\left(\varepsilon - \| \mathbf{s}_i - \mathbf{s}_j \|\right), \qquad H(x) = \begin{cases} 1, & x \geq 0 \\ 0, & x < 0 \end{cases} \tag{2}$$

where ε is the similarity threshold, $\| \cdot \|$ is a norm and $H(\cdot)$ is the Heaviside step function. In this study, m and τ are both set to 1, so each state is a single sample. The threshold step in Equation (2) is also omitted. Each image therefore holds the continuous distance $|x_i - x_j|$ instead of the binary value $R_{i,j}$. With m = 1, the image keeps the full time resolution of the short signal, and each pixel maps directly to a pair of time instants. The continuous distance also keeps fine changes in similarity that binarization would remove. These graded textures give the network more detail for learning repeated patterns at several scales. Because the transform compares every pair of time instants, small waveform differences between AW and DW signals, which are not always visually apparent in the raw time series, become separated texture patterns after encoding. Each matrix is resized to a 224 × 224 image for classification.

From a physical standpoint, the value of this encoding follows from the weld-pool dynamics: the melt pool and keyhole oscillate and briefly return toward a quasi-stable configuration many times within a single weld, so the emitted photodiode signal exhibits

partial recurrence instead of a smooth monotonic trend. It is this recurrence structure, more than the instantaneous signal amplitude, that a stable keyhole reproduces and an unstable or collapsing keyhole disrupts. This link between photoelectric radiation and keyhole/melt-pool behavior has recently been examined experimentally, providing physical support for using photoelectric signals, in addition to post-weld images, to characterize weld quality [43]. In a recurrence plot, continuous diagonal lines indicate periodic, repeating behavior, extended blocks indicate segments in which the process remains in similar states, and isolated points indicate abrupt changes.

Figure 2 shows representative signals and their recurrence plots. Acceptable welds produce a regular, fine-grained recurrent texture inside the main weld segment, whereas many defective welds produce large uniform blocks or an irregular pattern in which the fine texture is lost. Two features of these images shaped the network design in Section 3.4. First, the discriminative information lies at two scales at once: the coarse block layout of the plot and the fine periodic texture within it. Second, recurrence is a relation between distant time instants, so the relevant structure is spread over the whole image instead of being concentrated in a local object. Recurrence-plot encoding was compared with three alternative encodings on the infrared band, namely the Gramian angular difference field, the Gramian angular summation field and the Markov transition field, as reported in Section 4.2.

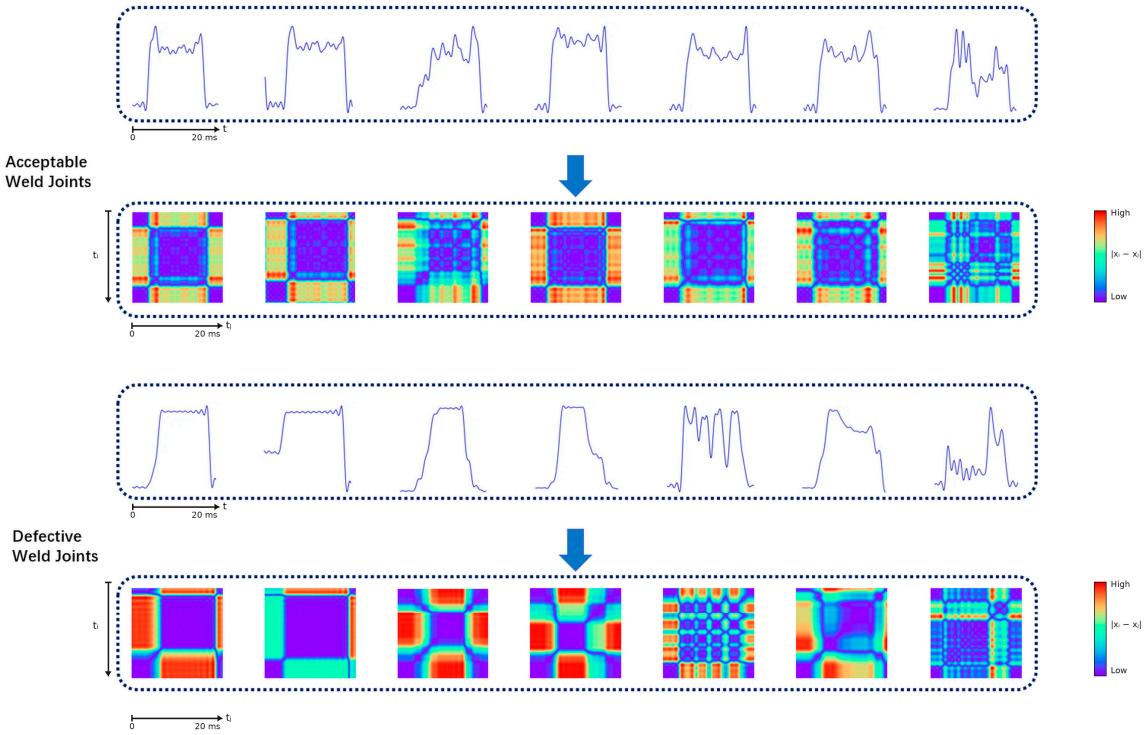


Figure 2. Raw photoelectric signals and their recurrence-plot encodings for acceptable weld joints (AW) and defective weld joints (DW). The time axis of each signal and both axes of each recurrence plot span about 20 ms (400 samples at 20 kHz). The color shows the distance $|x_i - x_j|$ from low (purple) to high (red).

3.4. BW-YOLO Network Architecture

Figure 3 shows the overall architecture of BW-YOLO. BW-YOLO uses a YOLO11-based backbone for whole-image classification [37]. As in the YOLO11n-cls baseline, the architecture does not require the detection neck used for object localization. Its absence is therefore not an architectural contribution relative to that baseline. The output of the final backbone stage is passed directly to a classification head composed of adaptive average pooling, dropout and a fully connected layer. After a 3 × 3 stride-2 stem convolution, Stage

1 applies a 3 × 3 stride-2 convolution and a C2f block to extract initial features from the recurrence-plot image. Stages 2 to 4 each combine a 3 × 3 stride-2 down-sampling convolution with a WT-ShuffleOSA block, and the final feature map is refined by a DCR-PSA block before classification.

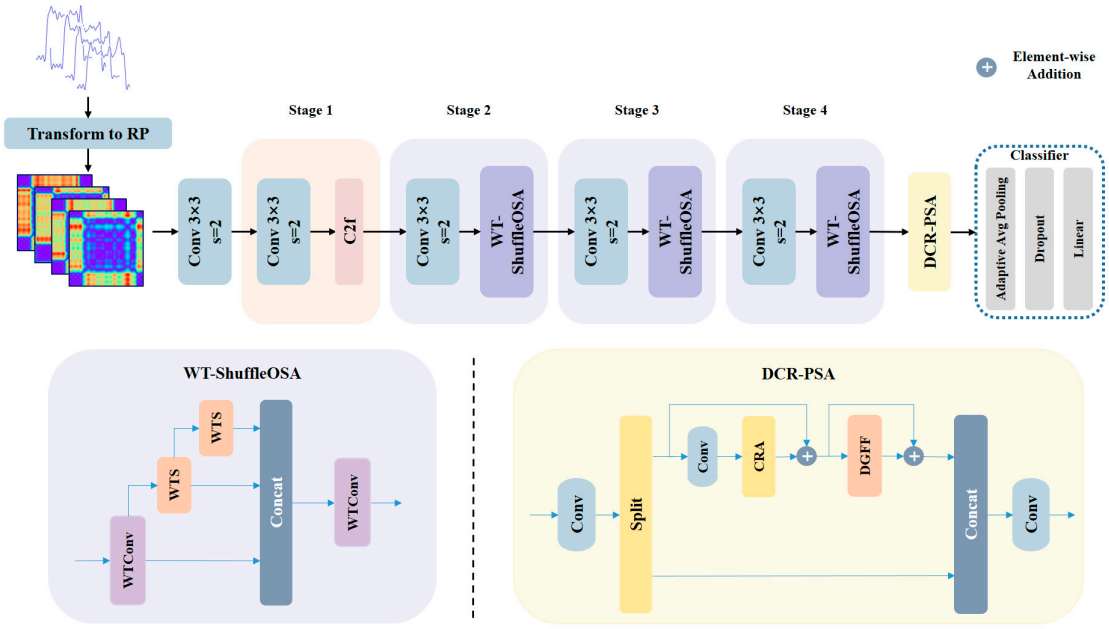


Figure 3. Overall architecture of BW-YOLO.

The choice of these modules follows the two features of recurrence-plot images noted in Section 3.3. The Haar decomposition in WT-ShuffleOSA splits each feature map into one low-frequency part and three high-frequency parts. The low-frequency part follows the coarse block layout of the plot. The high-frequency parts follow the fine recurrent texture. With this fixed split, the network does not have to learn the separation of the two scales from a small dataset. The channel reduction attention in DCR-PSA links distant positions of the plot, and these positions correspond to distant time instants of the weld. This attention runs at a cost close to that of the baseline attention layer. This proposed correspondence between image structure and network structure motivates the design; the experiments evaluate the resulting configurations without isolating the mechanisms responsible for their performance.

3.4.1. WT-ShuffleOSA Module

The original C3k2 block in YOLO11 relies on fixed-scale convolution kernels and multi-branch channel splitting, which makes it difficult to capture global context and local detail jointly and can leave redundant features unresolved in deeper layers. For recurrence-plot images, this matters because the coarse block layout and the fine recurrent texture must be modeled together. WT-ShuffleOSA (Figure 4) addresses this by combining wavelet transform convolution (WTConv) [44] with a channel shuffle operation [45] and the one-shot aggregation design used in RCS-OSA [46].

WTConv decomposes the input feature map with the Haar wavelet, implemented as four fixed 2 × 2 filters:

$$f_{LL} = 1/2\begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix}, \quad f_{LH} = 1/2\begin{bmatrix} 1 & -1 \\ 1 & -1 \end{bmatrix}, \quad f_{HL} = 1/2\begin{bmatrix} 1 & 1 \\ -1 & -1 \end{bmatrix}, \quad f_{HH} = 1/2\begin{bmatrix} 1 & -1 \\ -1 & 1 \end{bmatrix} \quad (3)$$

Applying these filters with stride 2 to an input $\mathbf{X} \in \mathbb{R}^{C\times H\times W}$ yields four sub-bands at half resolution:

$$[\mathbf{X}_{LL}, \mathbf{X}_{LH}, \mathbf{X}_{HL}, \mathbf{X}_{HH}] = \mathrm{Conv}([f_{LL}, f_{LH}, f_{HL}, f_{HH}], \mathbf{X}) \tag{4}$$

The low-frequency component $\mathbf{X}_{LL}$ carries the global structure of the feature map, and the high-frequency components $\mathbf{X}_{LH}$, $\mathbf{X}_{HL}$ and $\mathbf{X}_{HH}$ carry horizontal, vertical and diagonal detail. A small depth-wise convolution is applied to each sub-band, and the inverse wavelet transform (IWT) returns the result to the spatial domain:

$$\mathbf{Y} = \mathrm{Conv}(\mathbf{W}, [\mathbf{X}_{LL}, \mathbf{X}_{LH}, \mathbf{X}_{HL}, \mathbf{X}_{HH}]), \qquad \mathbf{X}' = \mathrm{IWT}(\mathbf{Y}) \tag{5}$$

where $\mathbf{W}$ denotes the convolution weights. As shown in Figure 4, the low-frequency component is decomposed a second time before reconstruction, and the wavelet branch is added to a spatial-domain convolution of the input. Because each level halves the resolution of the low-frequency component, a small kernel aggregates context over a receptive field that grows exponentially with the number of levels, while the number of parameters grows only linearly. For a recurrence-plot image, the low-frequency path therefore models the block layout of the plot, and the high-frequency paths respond to the fine diagonal and checkerboard texture that distinguishes stable from unstable welding.

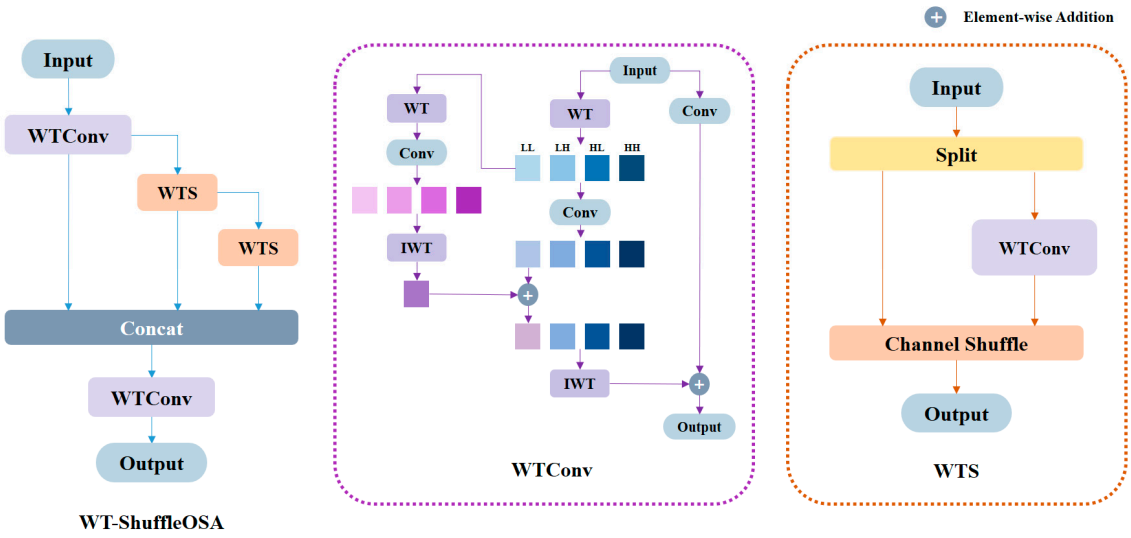


Figure 4. Structure of the WT-ShuffleOSA module, including the wavelet transform convolution (WTConv) and channel shuffle (WTS) units.

Channel shuffle exchanges information between channel groups without adding parameters. The input is reshaped into g groups, the group and channel dimensions are transposed, and the tensor is reshaped back:

$$\mathrm{Shuffle}(\mathbf{X}) = \mathrm{Reshape}(\mathrm{Transpose}(\mathrm{Reshape}(\mathbf{X}, [g, C/g, H, W])), [C, H, W]) \tag{6}$$

A WTS unit splits its input $\mathbf{Z}$ evenly along the channel dimension into $\mathbf{Z}_1$ and $\mathbf{Z}_2$, applies WTConv to $\mathbf{Z}_2$ only, and shuffles the concatenated result, so that wavelet-processed channels and untouched channels are mixed:

$$\mathrm{WTS}(\mathbf{Z}) = \mathrm{Shuffle}(\mathrm{Concat}(\mathbf{Z}_1, \mathrm{WTConv}(\mathbf{Z}_2))) \tag{7}$$

WT-ShuffleOSA applies one WTConv layer followed by channel shuffle, cascades two WTS units, and aggregates the three intermediate representations in a single concatenation, after which a final WTConv fuses them and restores the channel dimension:

$$\mathbf{X}_1 = \mathrm{Shuffle}(\mathrm{WTConv}(\mathbf{X})), \quad \mathbf{X}_2 = \mathrm{WTS}(\mathbf{X}_1), \quad \mathbf{X}_3 = \mathrm{WTS}(\mathbf{X}_2) \tag{8}$$

$$\mathbf{X}_{\text{out}} = \text{WTConv}(\text{Concat}(\mathbf{X}_1, \mathbf{X}_2, \mathbf{X}_3)) \tag{9}$$

The cascade enlarges the receptive field step by step, and the one-shot aggregation keeps features from every depth available to the output without the repeated concatenations of dense connections. Because shallow layers mainly encode pixel-level edges and corners, WT-ShuffleOSA is placed only in Stages 2 to 4, where deep, low-resolution feature maps benefit most from multiscale modeling at a modest computational cost.

3.4.2. DCR-PSA Module

The C2PSA block used in YOLO11 applies standard multi-head self-attention together with a simple feed-forward network, but the quadratic cost of global similarity computation becomes expensive on high-resolution feature maps, and the feed-forward network offers little cross-scale feature fusion. Because recurrence is a relation between distant time instants, global context remains important for recurrence-plot images, and DCR-PSA (Figure 5) retains it at a reduced cost. The input $\mathbf{X}$ is first remapped by a 1×1 convolution and split evenly along the channel dimension:

$$[\mathbf{A}, \mathbf{B}] = \text{Split}(\text{Conv}_{1\times1}(\mathbf{X})), \qquad \mathbf{A}, \mathbf{B} \in \mathbb{R}^{\frac{C}{2}\times H\times W} \tag{10}$$

Branch $\mathbf{A}$ is kept unchanged as a shortcut, and branch $\mathbf{B}$ is refined by a channel reduction attention (CRA) unit, adapted from the context modeling strategy of MetaSeg [47], and a dual-scale gated feed-forward (DGFF) unit, adapted from the histogram transformer [48], each with a residual connection:

$$\mathbf{B}_1 = \mathbf{B} + \text{CRA}(\text{Conv}(\mathbf{B})), \qquad \mathbf{B}_2 = \mathbf{B}_1 + \text{DGFF}(\mathbf{B}_1), \qquad \mathbf{Y} = \text{Conv}_{1\times1}(\text{Concat}(\mathbf{A}, \mathbf{B}_2)) \tag{11}$$

In CRA, the feature map is flattened into $N = HW$ tokens. The query is obtained by a channel-reduced linear projection, whereas the key and value are computed from an average-pooled copy of the tokens, $P(\cdot)$, which reduces the number of key and value tokens from N to N':

$$\mathbf{Q} = \mathbf{B}\mathbf{W}_Q, \quad \mathbf{K} = P(\mathbf{B})\mathbf{W}_K, \quad \mathbf{V} = P(\mathbf{B})\mathbf{W}_V, \qquad \text{CRA}(\mathbf{B}) = \text{Softmax}\left(\frac{\mathbf{Q}\mathbf{K}^\top}{\sqrt{d_r}}\right)\mathbf{V} \tag{12}$$

where $\mathbf{W}_Q, \mathbf{W}_K \in \mathbb{R}^{C'\times d_r}$ project to a reduced channel dimension d_r and $\mathbf{W}_V \in \mathbb{R}^{C'\times C'}$, with $C' = C/2$. The attention map has size $N \times N'$ instead of $N \times N$, and its similarity computation is carried out in the reduced dimension d_r, so global dependencies are modeled at a fraction of the cost of standard multi-head self-attention.

DGFF expands the channels with a 1×1 convolution and a pixel shuffle, splits the result into two parts, and processes them in parallel with a 5×5 depth-wise convolution, which captures coarse spatial context, and a 3×3 convolution with dilation rate 2, which captures finer texture over an enlarged field. The second branch, passed through a Mish activation, gates the first:

$$[\mathbf{Z}_1, \mathbf{Z}_2] = \text{Split}(\text{PS}(\text{Conv}_{1\times1}(\mathbf{B}_1))) \tag{13}$$

$$\text{DGFF}(\mathbf{B}_1) = \text{Conv}_{1\times1}(\text{PU}(\text{DWConv}_{5\times5}(\mathbf{Z}_1) \odot \text{Mish}(\text{DConv}_{3\times3}(\mathbf{Z}_2)))) \tag{14}$$

where PS and PU denote pixel shuffle and pixel unshuffle and $\odot$ is element-wise multiplication. The gate lets local structural responses select which coarse-scale features are passed on. By processing only half of the channels and leaving the other half as a shortcut, DCR-PSA models global dependencies and multiscale local detail jointly while adding a small computational overhead.

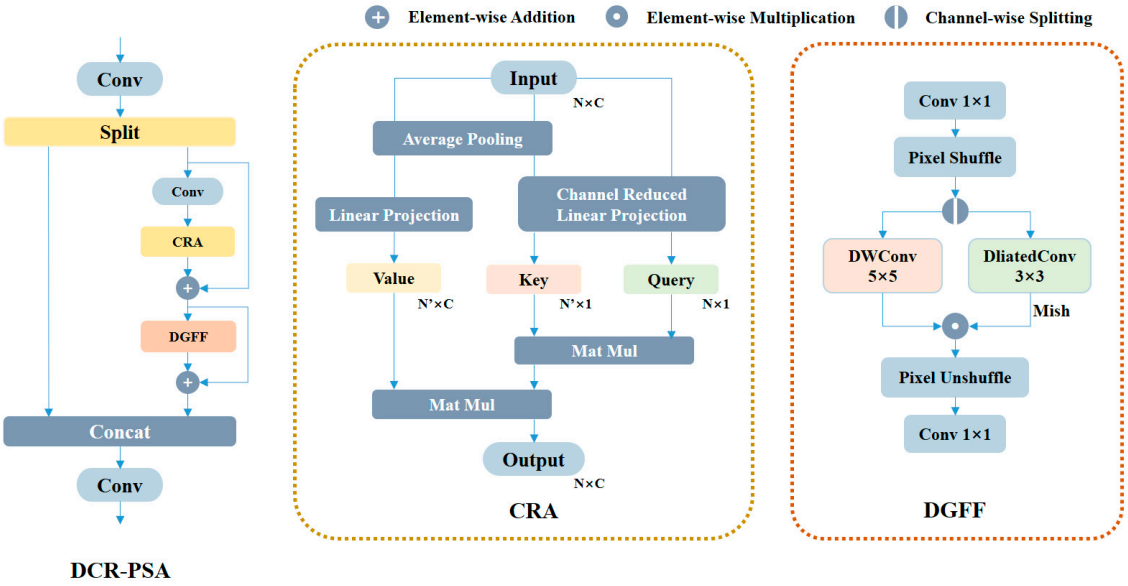


Figure 5. Structure of the DCR-PSA module, including the channel reduction attention (CRA) and dual-scale gated feed-forward (DGFF) units.

3.5. Training and Evaluation

BW-YOLO was implemented in PyTorch2.2.2 and trained on a single NVIDIA GeForce RTX 3090 GPU. Stochastic gradient descent was used with a base learning rate of 1×10^{-3}, a momentum of 0.937 and a weight decay of 0.0005, and the network was trained for 100 epochs with a batch size of 16 (Table 4). Categorical cross-entropy was used as the loss function, and a softmax layer produced the output class probabilities. All networks compared in Section 4 were trained with the same protocol on the same partition.

Table 4. Training configuration.

Training Parameter	Value
Optimizer	SGD
Base learning rate	1×10^{-3}
Weight decay	0.0005
Momentum	0.937
Batch size	16
Training epochs	100
Input image size	224 × 224

Recognition performance was assessed with accuracy, precision, recall and the F1-score:

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN}, \qquad \text{Precision} = \frac{TP}{TP + FP} \tag{15}$$

$$\text{Recall} = \frac{TP}{TP + FN}, \qquad F1 = \frac{2 \cdot \text{Precision} \cdot \text{Recall}}{\text{Precision} + \text{Recall}} \tag{16}$$

where TP, TN, FP and FN are the numbers of true positives, true negatives, false positives and false negatives on the validation subset. Because the validation subsets are balanced and of equal size, the number of misclassified joints, $FP + FN$, is also reported; its relative reduction expresses how many of a baseline's errors a method removes. Model complexity was assessed with the parameter count and the number of floating-point operations (GFLOPs), which are hardware-independent measures of inference cost and are reported in the same way for every network compared in Section 4.4.

4. Results

The experiments answer four questions in turn: what each proposed module contributes, whether the recurrence plot is the right encoding, whether the improvement holds across optical bands, and where BW-YOLO stands relative to established classification networks.

4.1. Contribution of Each Module

The contribution of each proposed module was evaluated on the infrared-band dataset using YOLO11n-cls as the baseline (Table 5). Each module improved all four recognition metrics on its own. WT-ShuffleOSA raised accuracy, F1-score, precision and recall by 1.67, 1.66, 2.53 and 0.84 percentage points while reducing the parameter count by 0.09 M and the computational cost from 3.3 to 2.9 GFLOPs, so it improved accuracy and efficiency in a single step. DCR-PSA raised the same metrics by 1.67, 1.72, 0.91 and 2.50 percentage points for an additional 0.06 M parameters and no change in GFLOPs, giving the largest gain in recall. The complete BW-YOLO network improved accuracy, F1-score, precision and recall by 2.09, 2.10, 2.55 and 1.67 percentage points over the baseline, while reducing the parameter count from 1.53 M to 1.50 M and the computational cost from 3.3 to 2.8 GFLOPs (−15%). In terms of errors, the number of misclassified validation joints fell from 13 to 8, a reduction of 38%. The ablation evaluates WT-ShuffleOSA and DCR-PSA as complete modules. Although their constituent operators were investigated in earlier studies [44–48], those evaluations do not establish their individual contributions within BW-YOLO. The present results therefore support comparison of the tested module configurations but do not identify which operators or interactions account for the observed differences.

Table 5. Ablation study on the infrared-band dataset (misclassified joints out of 240 validation samples).

Network	WT-ShuffleOSA	DCR-PSA	Acc (%)	F1 (%)	Prec. (%)	Rec. (%)	Params (M)	GFLOPs	Misclassified
YOLO11n-cls (baseline)			94.58	94.51	95.73	93.33	1.53	3.3	13
+ WT-ShuffleOSA	✓		96.25	96.17	98.26	94.17	1.44	2.9	9
+ DCR-PSA		✓	96.25	96.23	96.64	95.83	1.59	3.3	9
BW-YOLO (proposed)	✓	✓	**96.67**	**96.61**	**98.28**	95.00	1.50	**2.8**	**8**

4.2. Choice of Signal Encoding

Because the encoding strategy determines the information available to the classifier, four encoding methods, namely the Gramian angular difference field (GADF), the Gramian angular summation field (GASF), the Markov transition field (MTF) and the recurrence plot (RP), were compared on the infrared-band dataset using the YOLO11n-cls backbone (Table 6). RP achieved the highest accuracy (94.58%), F1-score (94.51%), precision (95.73%) and recall (93.33%). GADF and MTF both reached an accuracy of 93.75% and an F1-score of 93.62%, and GASF reached 91.25%. On this validation partition, RP produced 13 misclassifications, compared with 15 for GADF and MTF and 21 for GASF, and was therefore selected as the input representation for BW-YOLO. The two-sample advantage over GADF and MTF is small and, without repeated evaluation, does not establish a reliable performance difference. Section 5.1 discusses possible reasons for the suitability of this representation.

Table 6. Comparison of signal-to-image encoding methods on the infrared-band dataset (YOLO11n-cls backbone).

Encoding	Acc (%)	F1 (%)	Precision (%)	Recall (%)
GADF	93.75	93.62	95.65	91.67
GASF	91.25	91.06	93.04	89.17
MTF	93.75	93.62	95.65	91.67
RP (used)	94.58	94.51	95.73	93.33

4.3. Consistency Across Optical Bands

BW-YOLO was compared with the YOLO11n-cls baseline on all three single-band datasets to assess whether the observed improvement was consistent across optical bands (Table 7). The three channels reflect different emission mechanisms (Section 3.1), but their signals were recorded from the same joints and therefore provide related evaluations rather than independent replications. BW-YOLO improved accuracy by 2.09 percentage points on the infrared band (94.58% to 96.67%), by 1.25 percentage points on the visible band (95.83% to 97.08%) and by 0.83 percentage points on the reflected band (97.50% to 98.33%). The corresponding numbers of misclassified joints decreased from 13 to 8, from 10 to 7 and from 6 to 4, respectively, giving relative error reductions of approximately 38%, 30% and 33%. These counts refer to 240 validation joints evaluated separately in each band, not to 720 distinct joints. With 240 validation samples per band, one sample equals 0.42 percentage points of accuracy. On the infrared band, the 95% Wilson confidence intervals of the accuracy are 91.0–96.8% for YOLO11n-cls and 93.6–98.3% for BW-YOLO. These intervals describe uncertainty in the individual accuracy estimates and do not establish the statistical significance of the difference between models. All four metrics increased in each band on this fixed partition, although they are derived from the same predictions and should not be interpreted as independent evidence of improvement. The results show a consistent descriptive advantage for BW-YOLO under the evaluated conditions; its robustness across training runs and alternative partitions remains to be assessed.

Table 7. Performance of BW-YOLO and the YOLO11n-cls baseline across the three optical bands (misclassified joints out of 240 validation samples per band; error reduction relative to the baseline).

Optical Band	Network	Acc (%)	F1 (%)	Precision (%)	Recall (%)	Misclassified	Error Reduction
Infrared	YOLO11n-cls	94.58	94.51	95.73	93.33	13	Reference
	BW-YOLO	96.67	96.61	98.28	95.00	8	38%
Visible	YOLO11n-cls	95.83	95.97	92.97	99.17	10	Reference
	BW-YOLO	97.08	97.17	94.49	100.00	7	30%
Reflected	YOLO11n-cls	97.50	97.46	99.14	95.83	6	Reference
	BW-YOLO	98.33	98.32	99.15	97.50	4	33%

On the visible band, BW-YOLO reached a recall of 100%. The reflected band gave the highest absolute accuracy for both networks, and BW-YOLO reached 98.33% on it.

4.4. Comparison with Representative Classification Networks

BW-YOLO was further compared with nine representative classification networks on the infrared-band dataset: two heavyweight networks, VGG16 and VGG19 [49]; two mid-sized networks, ResNet50 [50] and ResNeXt50 [51]; and five lightweight or efficiency-oriented networks, MobileNetV3 [52], ShuffleNetV2 [45], EfficientNetV2 [53], DenseNet121 [54] and ConvNeXt [55], each trained under the same protocol (Table 8). BW-YOLO achieved the highest accuracy (96.67%), F1-score (96.61%) and precision (98.28%) of the ten networks.

Table 8. Comparison of BW-YOLO with representative classification networks on the infrared-band dataset.

Class	Network	Acc (%)	F1 (%)	Precision (%)	Recall (%)	Params (M)	GFLOPs
Heavyweight	VGG16	95.00	95.04	94.26	95.83	134.27	15.47
	VGG19	95.83	95.90	94.35	97.50	139.58	19.63
Mid-sized	ResNet50	96.25	96.27	95.87	96.67	23.51	4.13
	ResNeXt50	95.00	95.04	94.26	95.83	22.98	4.29
Efficiency-oriented	MobileNetV3	95.42	95.44	95.04	95.83	1.52	0.06
	ShuffleNetV2	96.25	96.20	97.44	95.00	1.26	0.15
	EfficientNetV2	95.42	95.40	95.80	95.00	20.18	2.90
	DenseNet121	95.83	95.80	96.61	95.00	6.96	2.90
	ConvNeXt	92.08	91.77	95.50	88.33	87.55	15.37
Proposed	BW-YOLO	96.67	96.61	98.28	95.00	1.50	2.80

Figure 6 places the ten networks in the accuracy–parameter plane. BW-YOLO sits at the upper-left corner: no other network is both smaller and more accurate. Against the heavyweight networks, it is more accurate than VGG16 and VGG19 with 89 and 93 times fewer parameters. Against the mid-sized networks, it exceeds ResNet50 in accuracy with one-sixteenth of the parameters and one-third fewer GFLOPs. At a matched computational budget, BW-YOLO (2.8 GFLOPs) is more accurate than EfficientNetV2 and DenseNet121 (2.9 GFLOPs each) by 1.25 and 0.84 percentage points while using 13.5 and 4.6 times fewer parameters. Within the sub-2 M-parameter class, it achieves the highest accuracy, F1-score and precision, exceeding ShuffleNetV2 by 0.84 points in precision and MobileNetV3 by 1.25 points in accuracy. MobileNetV3 and ShuffleNetV2 were designed for minimum FLOPs on mobile processors, and their 0.06 and 0.15 GFLOPs reflect that design target; BW-YOLO is designed for the highest recognition performance at a parameter count of the same order, which is the quantity that governs model storage and memory traffic on industrial edge hardware.

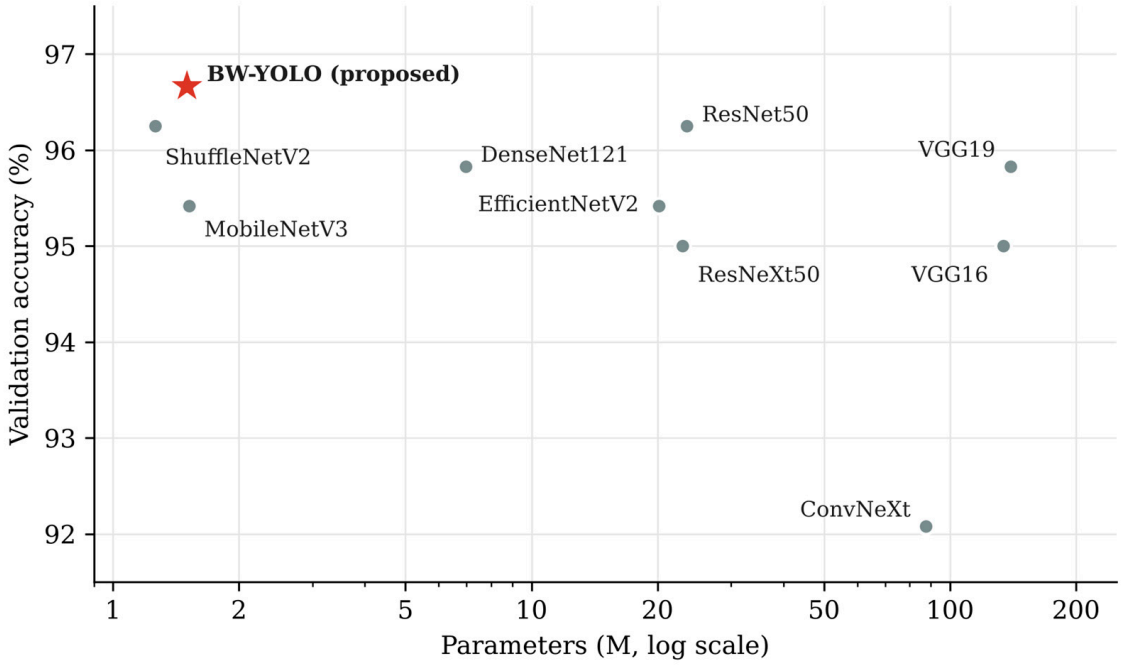


Figure 6. Validation accuracy versus parameter count of the ten networks on the infrared-band dataset (data from Table 8). The parameter axis is logarithmic.

5. Discussion

5.1. Interpretation of the Encoding and Architecture Results

The module-level ablation suggests different effects on the balance between precision and recall. WT-ShuffleOSA yields the larger gain in precision, whereas DCR-PSA yields the larger gain in recall. These observations are compatible with the intended roles of multiscale feature extraction and global context modelling, respectively, but do not establish those mechanisms as the causes of the improvements. The complete model achieved the highest accuracy and precision among the tested configurations, while its recall was lower than that of the DCR-PSA-only configuration. Its accuracy exceeded either single-module configuration by one correctly classified validation joint. Consequently, the results do not establish a robust synergistic effect between the modules. Component-level ablations and repeated training would be needed to distinguish their individual effects and interactions.

The observed performance of RP motivates consideration of how its representation may suit the short, non-stationary photodiode signals recorded during battery laser welding. RP keeps the continuous distance between every pair of samples. It does not map amplitudes to angles as the Gramian fields do, and it does not quantize amplitudes as the Markov transition field does. These properties provide a rationale for investigating RP as a representation of temporal similarity in welding signals. The reported relationship between photoelectric radiation and keyhole/melt-pool behaviour [43] also provides a physical motivation for this approach. However, the present encoding comparison does not establish that recurrence structure is the specific source of the classification advantage, nor that RP preserves more task-relevant information than the alternative encodings. The proposed correspondence between RP texture and the network architecture should therefore be regarded as a design rationale whose contribution requires further experimental isolation.

5.2. Optical Band Selection for Inline Monitoring

The higher absolute accuracy obtained on the reflected-laser band, for both YOLO11n-cls and BW-YOLO, is consistent with the physical origin of that signal. The reflected component tracks the opening, stability and collapse of the keyhole directly (Section 3.1), whereas the thermal radiation captured in the infrared band and the plasma and spatter luminescence captured in the visible band are more strongly affected by transient fluctuations in the weld pool and plume. The reflected-laser band gave the highest observed accuracy for both networks on the evaluated partition, with BW-YOLO reaching 98.33% accuracy and 99.15% precision. This result supports investigating that band for single-channel monitoring, although its relative performance under other production conditions remains to be established.

The infrared band, the most demanding of the three, is where the architecture adds the most, removing 38% of the baseline's errors. BW-YOLO improved all four reported metrics relative to the baseline in each band on this partition. This consistency is encouraging, but the shared joints and absence of repeated training limit conclusions about robustness across optical modalities. Furthermore, because the module-level ablation was restricted to the infrared band, the separate contributions of WT-ShuffleOSA and DCR-PSA in the other bands were not assessed.

5.3. Implications for Deployment

The comparison with representative classification networks places BW-YOLO on the favorable side of the accuracy–efficiency trade-off (Figure 6). With 1.50 M parameters, the model occupies about 6 MB in 32-bit floating point, small enough to be stored and executed alongside other inspection software on an industrial PC or an embedded module. For a 400-sample signal, constructing the continuous distance matrix involves

$400 \times 400 = 1.6 \times 10^5$ pairwise differences and their absolute values. This arithmetic count covers only the distance calculation and does not quantify the full cost of preprocessing, image conversion, resizing and data transfer. End-to-end latency was not measured, so the encoding overhead and compatibility with a specific welding cycle time remain to be established. These measurements are necessary to assess deployment feasibility alongside the reported recognition performance and model complexity.

Because the photodiode signals are acquired during welding, the proposed approach could support screening of individual joints and referral of suspect joints for further inspection. Its suitability for production-line use would require validation of latency and classification performance under operational conditions, including the actual defect prevalence and the consequences of missed defects and false alarms [7,15].

5.4. Scope and Extensions

The present study targets the pass/fail decision that governs inline inspection, on a single welding platform and material combination.

The evaluation used one fixed training–validation partition, without repeated training across random seeds or alternative partitions. Consequently, variability associated with initialization and sample selection remains unquantified. The Wilson confidence intervals describe uncertainty in validation accuracy under a binomial model; they do not capture training variability or establish the statistical significance of differences between models. The three optical bands were recorded from the same joints and therefore do not constitute independent replications. Furthermore, the same validation partition was used to compare encodings and architectural variants, without a separate test set for evaluating the selected configuration.

The encoding comparison was restricted to the infrared band, where RP produced two fewer misclassifications than GADF and MTF among 240 validation samples. This result supported its selection for the present study but does not establish superiority across optical bands or datasets. The module-level ablation assesses WT-ShuffleOSA and DCR-PSA as complete units and does not isolate the contributions of their constituent operators. Direct time-series classifiers were not evaluated, so the relative accuracy and computational efficiency of the proposed pipeline compared with 1D-CNNs or temporal convolutional networks remain undetermined.

The balanced dataset does not reproduce the defect prevalence of the production line, and precision under operational conditions may differ from the reported values. Parameter counts and operation counts characterize model complexity, but end-to-end latency, including preprocessing and RP generation, was not measured on deployment hardware. Compliance with a specific production cycle time therefore remains to be established.

Four directions for further work follow from these findings and limitations. First, since the three optical bands carry complementary physical information and were evaluated here separately, fusing them, following strategies demonstrated for other multi-sensor welding and cladding processes [28,56], offers a route to further accuracy without additional sensing hardware. Second, the same encoding and network can be extended from binary classification to the identification of specific defect types and to further materials and process windows. Third, measuring inference latency on the industrial-PC and embedded hardware used on battery lines, including after compression or quantization, will translate the parameter and GFLOPs advantages reported here into cycle-time figures for a specific production line. Fourth, further validation should include repeated training across random seeds and joint-level partitions, evaluation on an independent test set, encoding comparisons

across all three bands, component-level ablations, and comparisons with one-dimensional time-series models such as 1D-CNNs and temporal convolutional networks.

6. Conclusions

This study presented BW-YOLO, a lightweight network for identifying acceptable and defective welds in lithium-ion battery laser welding from multi-band photoelectric signals, built on the principle that the signal representation and the classification network should be designed together. Recurrence-plot encoding converts visible-light, infrared and laser-reflection signals into images that preserve the recurrence structure of the keyhole and melt-pool dynamics, and on the infrared band it gave 0.83 percentage points higher accuracy than the Gramian angular difference field and the Markov transition field and 3.33 percentage points higher accuracy than the Gramian angular summation field. The WT-ShuffleOSA and DCR-PSA modules were integrated into a YOLO11-based classification architecture to capture multiscale texture and long-range relationships in the encoded images. Each module individually improved all four recognition metrics relative to the baseline on the evaluated infrared validation partition.

On the fixed validation partition, BW-YOLO reduced the baseline's misclassifications by 30–38% across the three optical bands, with a reported reduction in operation count of approximately 15%. These results describe the observed accuracy–complexity trade-off under the evaluated conditions.

On the infrared band it reached 96.67% accuracy, 96.61% F1-score and 98.28% precision with 1.50 M parameters and 2.8 GFLOPs, giving the highest observed accuracy among the ten networks compared on that partition. On the reflected-laser band it reached 98.33% accuracy, the highest accuracy observed among the three individual optical bands on the evaluated validation partition.

The findings support further investigation of recurrence-plot encoding combined with a compact classification architecture for weld-quality monitoring. The proposed correspondence between signal representation and network structure provides a design rationale, while its advantage over alternative approaches requires confirmation through repeated evaluation and direct time-series comparisons.

The approach is therefore a candidate for photodiode-based screening of individual welds, subject to validation of end-to-end latency and classification performance under production conditions.

Author Contributions: Conceptualization, N.B.; methodology, J.L.; software, J.L.; validation, Y.F.; formal analysis, Y.F.; investigation, J.L. and R.B.; resources, A.S.; data curation, J.L.; writing—original draft preparation, J.L.; writing—review and editing, N.B., Y.F., A.S. and R.B.; supervision, N.B. and A.S.; project administration, N.B. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported by the 2026 Dongguan Outstanding Engineer Joint Training Key Project, under the project "Research and Application of Key Technologies for End-to-End Integration of Industrial Embodied Intelligent Agents" (Grant No. 20261501200132).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The data presented in this study were collected on an industrial production line and are subject to the confidentiality requirements of the industrial partner. The data are therefore not publicly available.

Acknowledgments: During the preparation of this manuscript, the authors used Claude (Anthropic, Claude Opus 5.5) and ChatGPT T-6 Astra(Open AI) for the purposes of literature identification,

citation verification and language editing. The authors have reviewed and edited the output and take full responsibility for the content of this publication.

Conflicts of Interest: Author Jiaqi Li is employed by Liqun Automation Technology Co., Ltd. The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as potential conflicts of interest.

References

1. International Energy Agency (IEA). *Global EV Outlook 2026: Electric Vehicle Batteries*; IEA: Paris, France, 2026. Available online: https://www.iea.org/reports/global-ev-outlook-2026/electric-vehicle-batteries (accessed on 24 July 2026).
2. Wang, Q.; Jiang, B.; Li, B.; Yan, Y. A critical review of thermal management models and solutions of lithium-ion batteries for the development of pure electric vehicles. *Renew. Sustain. Energy Rev.* **2016**, *64*, 106–128. [CrossRef]
3. Chianese, G.; Franciosa, P.; Nolte, J.; Ceglarek, D.; Patalano, S. Characterization of photodiodes for detection of variations in part-to-part gap and weld penetration depth during remote laser welding of copper-to-steel battery tab connectors. *J. Manuf. Sci. Eng.* **2022**, *144*, 071004. [CrossRef]
4. Choi, S.-H.; Kim, J.-H.; Choi, H.-W. Ring beam modulation-assisted laser welding on dissimilar materials for automotive battery. *J. Manuf. Mater. Process.* **2025**, *9*, 28. [CrossRef]
5. Wang, M.; Zhang, L.; Liu, Y.; Hu, P.; Peng, X.; Xie, Z.; Ke, W. Machine learning in laser welding of electric vehicle batteries: Process control and quality assurance. *Mater. Des.* **2026**, *267*, 116205. [CrossRef]
6. Yang, Y.; Yang, R.; Pan, L.; Ma, J.; Zhu, Y.; Diao, T.; Zhang, L. A lightweight deep learning algorithm for inspection of laser welding defects on safety vent of power battery. *Comput. Ind.* **2020**, *123*, 103306. [CrossRef]
7. Schnell, J.; Reinhart, G. Quality management for battery production: A quality gate concept. *Procedia CIRP* **2016**, *57*, 568–573. [CrossRef]
8. Grabmann, S.; Harst, F.; Bernauer, C.; Weiss, T.; Zaeh, M.F. Analysis of photodiode signals for monitoring the laser beam welding process of cell-internal contacts in lithium-ion batteries. *Opt. Laser Technol.* **2024**, *175*, 110730. [CrossRef]
9. Chianese, G.; Franciosa, P.; Sun, T.; Ceglarek, D.; Patalano, S. Using photodiodes and supervised machine learning for automatic classification of weld defects in laser welding of thin foils copper-to-steel battery tabs. *J. Laser Appl.* **2022**, *34*, 042040. [CrossRef]
10. Cai, W.; Jiang, P.; Shu, L.; Geng, S.; Zhou, Q. Real-time monitoring of laser keyhole welding penetration state based on deep belief network. *J. Manuf. Process.* **2021**, *72*, 203–214. [CrossRef]
11. Cheng, Y.; Yu, R.; Zhou, Q.; Chen, H.; Yuan, W.; Zhang, Y.M. Real-time sensing of gas metal arc welding process—A literature review and analysis. *J. Manuf. Process.* **2021**, *70*, 452–469. [CrossRef]
12. Meyer, K.; Mahalec, V. Anomaly detection methods for infrequent failures in resistive steel welding. *J. Manuf. Process.* **2022**, *75*, 497–513. [CrossRef]
13. Xu, L.; Dong, S.; Wei, H.; Ren, Q.; Huang, J.; Liu, J. Defect signal intelligent recognition of weld radiographs based on YOLO V5-IMPROVEMENT. *J. Manuf. Process.* **2023**, *99*, 373–381. [CrossRef]
14. Yu, R.; Cao, Y.; Chen, H.; Ye, Q.; Zhang, Y.M. Deep learning based real-time and in-situ monitoring of weld penetration: Where we are and what are needed revolutionary solutions? *J. Manuf. Process.* **2023**, *93*, 15–46. [CrossRef]
15. Zhang, M.; Xu, X.; Polzer, J.; Liu, Q.; Chen, X.; Chen, X. Data-driven artificial intelligence methods for real-time welding defect diagnosis: A critical review and future outlook. *J. Manuf. Process.* **2026**, *161*, 49–63. [CrossRef]
16. Wang, J.; Ma, Y.; Zhang, L.; Gao, R.X.; Wu, D. Deep learning for smart manufacturing: Methods and applications. *J. Manuf. Syst.* **2018**, *48*, 144–156. [CrossRef]
17. Ancona, A.; Spagnolo, V.; Lugarà, P.M.; Ferrara, M. Optical sensor for real-time monitoring of CO_2 laser welding process. *Appl. Opt.* **2001**, *40*, 6019–6025. [CrossRef] [PubMed]
18. Kong, F.; Ma, J.; Carlson, B.; Kovacevic, R. Real-time monitoring of laser welding of galvanized high strength steel in lap joint configuration. *Opt. Laser Technol.* **2012**, *44*, 2186–2196. [CrossRef]
19. Andreev, A. Smart laser welding heads provide excellent quality. *Laser Tech. J.* **2009**, *5*, 20–22. [CrossRef]
20. Park, Y.W.; Park, H.; Rhee, S.; Kang, M. Real time estimation of CO_2 laser weld quality for automotive industry. *Opt. Laser Technol.* **2002**, *34*, 135–142. [CrossRef]
21. Ozkat, E.C. Photodiode signal patterns: Unsupervised learning for laser weld defect analysis. *Processes* **2025**, *13*, 121. [CrossRef]
22. Sotubadi, S.V.; Hosseinzadeh, A.; Ghassemi-Armaki, H.; Pour, M.M.; Ma, J.; Zhao, X.; Bracey, J.T.; Carlson, B.E.; Wang, P.E. Harnessing unlabeled plant data and labeled lab data for enhanced quality prediction in laser welding. *J. Manuf. Process.* **2026**, *157*, 1015–1034. [CrossRef]
23. Nazir, Q.; Shao, C. Online tool condition monitoring for ultrasonic metal welding via sensor fusion and machine learning. *J. Manuf. Process.* **2021**, *62*, 806–816. [CrossRef]

24. Schwarz, E.B.; Bleier, F.; Guenter, F.; Mikut, R.; Bergmann, J.P. Improving process monitoring of ultrasonic metal welding using classical machine learning methods and process-informed time series evaluation. *J. Manuf. Process.* **2022**, *77*, 54–62. [CrossRef]
25. Scott, R.; Stocco, D.; Sarafan, S.; Behnen, L.; Chertov, A.M.; Wanjara, P.; Maev, R.G. Automation of ultrasonic monitoring for resistance spot welding using deep learning. *J. Manuf. Mater. Process.* **2026**, *10*, 101. [CrossRef]
26. Cao, L.; Li, J.; Zhang, L.; Luo, S.; Li, M.; Huang, X. Cross-attention-based multi-sensing signals fusion for penetration state monitoring during laser welding of aluminum alloy. *Knowl.-Based Syst.* **2023**, *261*, 110212. [CrossRef]
27. Geiger, C.; Garkusha, P.; Kasper, B.; Mayerhofer, R.; Fuehrer, A.; Seitz, D.; Weiss, T.; Zaeh, M.F. Photodiode-based process monitoring for the ultrashort-pulsed laser structuring of the diffusion media for fuel cells. *J. Manuf. Process.* **2024**, *131*, 2016–2028. [CrossRef]
28. Chen, X.; Yao, M.; Kong, F.; Fu, Y.; Wu, J.; Zhang, H. In-situ quality monitoring of laser hot wire cladding process based on multi-sensing diagnosis and machine learning model. *J. Manuf. Process.* **2023**, *87*, 183–198. [CrossRef]
29. Wang, J.; Zhang, Z.; Zhang, S.; Qin, H.; Qin, R.; Huang, J.; Wen, G.; Chen, X. A novel wavelet convolutional network for monitoring interfacial bonding quality in LDED using AE signals. *J. Manuf. Process.* **2025**, *153*, 899–915. [CrossRef]
30. Eckmann, J.-P.; Kamphorst, S.O.; Ruelle, D. Recurrence plots of dynamical systems. *Europhys. Lett.* **1987**, *4*, 973–977. [CrossRef]
31. Hatami, N.; Gavet, Y.; Debayle, J. Classification of time-series images using deep convolutional neural networks. In *Proceedings of the Tenth International Conference on Machine Vision (ICMV 2017)*; SPIE: Bellingham, WA, USA, 2018; Volume 10696, pp. 242–249.
32. Wang, Z.; Oates, T. Encoding time series as images for visual inspection and classification using tiled convolutional neural networks. In *Proceedings of the Workshops at the 29th AAAI Conference on Artificial Intelligence, Austin, TX, USA, 25–30 January 2015*; Association for the Advancement of Artificial Intelligence (AAAI): Washington, DC, USA, 2015.
33. Abidi, A.; Ienco, D.; Ben Abbes, A.; Farah, I.R. Orthrus: Multi-scale land cover mapping from satellite image time series via 2D encoding and convolutional neural network. *Neural Comput. Appl.* **2024**, *36*, 19247–19265. [CrossRef]
34. Liu, Y.; Yuan, K.; Li, T.; Li, S.; Ren, Y. NDT method for line laser welding based on deep learning and one-dimensional time-series data. *Appl. Sci.* **2022**, *12*, 7837. [CrossRef]
35. Shi, H.; Gao, R.; Zhang, C.; Cao, Y.; Xu, Y.; Jin, L.; Zhao, P. Product quality prediction for injection molding based on temperature field infrared thermography and convolutional neural network. *J. Manuf. Process.* **2024**, *125*, 11–24. [CrossRef]
36. He, X.; Jin, J.; Chen, D.; Feng, Y. An integrated defect detection method based on context encoder and perception-enhanced aggregation for cylinder bores. *J. Manuf. Process.* **2024**, *114*, 196–212. [CrossRef]
37. Ultralytics. YOLO11. 2024. Available online: https://docs.ultralytics.com/ (accessed on 1 October 2024).
38. Liang, Y.; Feng, S.; Zhang, Y.; Xue, F.; Shen, F.; Guo, J. A stable diffusion enhanced YOLOV5 model for metal stamped part defect detection based on improved network structure. *J. Manuf. Process.* **2024**, *111*, 21–31. [CrossRef]
39. Sun, H.; Xia, L.; Zhou, Y.; Qian, W.; Wang, W.; Zhang, K. Online detection and evaluation of weld surface defects based on lightweight network VGG16-UNet and laser scanning. *J. Manuf. Process.* **2024**, *129*, 292–306. [CrossRef]
40. Zhang, L.; He, Y.; Diao, T.; Liu, Z. A real-time lightweight laser welding defect inspection algorithm based on deep learning. *Comput. Ind.* **2026**, *174*, 104413. [CrossRef]
41. 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]
42. Xu, J.; Liu, Q.; Xu, Y.; Xiao, R.; Hou, Z.; Chen, S. Review on the application of the attention mechanism in sensing information processing for dynamic welding processes. *J. Manuf. Mater. Process.* **2024**, *8*, 22. [CrossRef]
43. Zeng, D.; Wu, D.; Zhang, P.; Shi, H.; Lu, Q.; Yu, Z. Shedding light on correlation mechanism between the keyhole/melt pool behaviors and photoelectric radiation information during laser welding process. *Opt. Laser Technol.* **2025**, *192*, 113465. [CrossRef]
44. Finder, S.E.; Amoyal, R.; Treister, E.; Freifeld, O. Wavelet convolutions for large receptive fields. In *Computer Vision—ECCV 2024*; Lecture Notes in Computer Science; Springer: Cham, Switzerland, 2025; Volume 15112, pp. 363–380.
45. Ma, N.; Zhang, X.; Zheng, H.-T.; Sun, J. ShuffleNet V2: Practical guidelines for efficient CNN architecture design. In *Proceedings of the European Conference on Computer Vision (ECCV), Munich, Germany, 8–14 September 2018*; Springer: Cham, Switzerland, 2018; pp. 116–131.
46. Kang, M.; Ting, C.-M.; Ting, F.F.; Phan, R.C.-W. RCS-YOLO: A fast and high-accuracy object detector for brain tumor detection. In *Medical Image Computing and Computer Assisted Intervention—MICCAI 2023*; Springer: Cham, Switzerland, 2023; pp. 600–610.
47. Kang, B.; Moon, S.; Cho, Y.; Yu, H.; Kang, S. MetaSeg: MetaFormer-based global contexts-aware network for efficient semantic segmentation. In *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision (WACV), Waikoloa, HI, USA, 3–8 January 2024*; IEEE: Piscataway, NJ, USA, 2024; pp. 433–442.
48. Sun, S.; Ren, W.; Gao, X.; Wang, R.; Cao, X. Restoring images in adverse weather conditions via histogram transformer. In *Computer Vision—ECCV 2024*; Lecture Notes in Computer Science; Springer: Cham, Switzerland, 2025; Volume 15080, pp. 111–129.
49. Simonyan, K.; Zisserman, A. Very deep convolutional networks for large-scale image recognition. *arXiv* **2014**, arXiv:1409.1556.
50. He, K.; Zhang, X.; Ren, S.; Sun, J. Deep residual learning for image recognition. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA, 27–30 June 2016*; IEEE: Piscataway, NJ, USA, 2016; pp. 770–778.

51. Xie, S.; Girshick, R.; Dollár, P.; Tu, Z.; He, K. Aggregated residual transformations for deep neural networks. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, HI, USA, 21–26 July 2017*; IEEE: Piscataway, NJ, USA, 2017; pp. 1492–1500.
52. Howard, A.; Sandler, M.; Chu, G.; Chen, L.-C.; Chen, B.; Tan, M.; Wang, W.; Zhu, Y.; Pang, R.; Vasudevan, V.; et al. Searching for MobileNetV3. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV), Seoul, Republic of Korea, 27 October–2 November 2019*; IEEE: Piscataway, NJ, USA, 2019; pp. 1314–1324.
53. Tan, M.; Le, Q. EfficientNetV2: Smaller models and faster training. In *Proceedings of the 38th International Conference on Machine Learning (ICML)*; PMLR: Brookline, MA, USA, 2021; pp. 10096–10106.
54. Huang, G.; Liu, Z.; van der Maaten, L.; Weinberger, K.Q. Densely connected convolutional networks. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, HI, USA, 21–26 July 2017*; IEEE: Piscataway, NJ, USA, 2017; pp. 4700–4708.
55. Liu, Z.; Mao, H.; Wu, C.-Y.; Feichtenhofer, C.; Darrell, T.; Xie, S. A ConvNet for the 2020s. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), New Orleans, LA, USA, 18–24 June 2022*; IEEE: Piscataway, NJ, USA, 2022; pp. 11976–11986.
56. Chen, X.; Kong, F. Applying multi-sensor fusion strategy for enhanced predictions of clad hardness and dilution in laser hot-wire cladding process by an ensemble machine learning model. *J. Manuf. Process.* **2025**, *133*, 211–225. [CrossRef]

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.