Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Instance resolved measurement of corrosion pits in an aluminium lithium alloy using confocal topography

Hongsheng Liu
Rocket Force University of Engineering
Dejun Liu
liudjwxx801@yeah.net

Rocket Force University of Engineering
Gan Tian
Rocket Force University of Engineering
Xinzhi Yang
Rocket Force University of Engineering
Biyun Ren
Rocket Force University of Engineering

Article

Keywords:

Posted Date: November 11th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-11497786/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

Instance resolved measurement of corrosion pits in an aluminium lithium alloy using confocal topography

Hongsheng Liu, Dejun Liu*, Gan Tian, Xinzhi Yang and Biyun Ren

College of Missile Engineering, Rocket Force University of Engineering, Hongqing Street, Baqiao District, Xi'an, China.

*Correspondence: Dejun Liu (liudjwxx801@yeah.net).

Abstract

Aluminium-lithium alloys support lightweight aerospace structures, but pitting creates localized material loss and potential fatigue-crack initiation sites. Quantifying this damage requires individual pit boundaries linked to physical depth. Here we develop PitRPM to connect instance recognition with geometric measurement in confocal height fields of 2195-T8 alloy. Height, local depth residual and gradient channels feed a mask segmenter and an optional vision-language proposal branch. Accepted masks are measured on the original height field. Across 120 fields containing 1,138 reviewed pits, field-level validation yielded an F1 score of 0.826 and mean mask intersection-over-union of 0.758. Matched-pit relative errors were 31.8% for projected area, 0.8% for maximum depth and 5.2% for mean depth. Additional proposals increased recall with a corresponding reduction in precision. The framework connects automated pit recognition to physical descriptors and distinguishes the boundary sensitivity of area from the pit-bottom sensitivity of depth.

Introduction

Aluminium-lithium alloys are used in lightweight aerospace structures because lithium additions reduce density while supporting high specific mechanical performance [1]. Exposure to moisture and aggressive species makes localized corrosion important for these applications. Concentrated material loss can penetrate the surface and provide sites for fatigue-crack initiation [2-4]. Individual pit geometry and interactions between neighbouring defects can influence the subsequent mechanical response [5-6]. The significance of this damage therefore depends on its spatial distribution, not simply the visibly affected surface fraction. A broad, shallow depression and a narrow, deep pit can occupy similar projected areas while representing different geometrical disturbances. Surface coverage alone removes this distinction and combines neighbouring features into one measure. An instance-resolved description instead links each footprint to its penetration and spatial position. Measuring pit location, projected area and depth provides the geometrical information needed to distinguish individual defects within a corroded field.

In precipitation-strengthened aluminium alloys, corrosion morphology reflects interactions between the matrix, intermetallic particles and the environment. Studies of Al-Cu-Li alloys have examined particle-associated attack using three-dimensional chemical imaging and microstructural characterization [7-8]. Processing and exposure effects have also been investigated in 2195 welded

joints, Sc-modified alloys and chloride-exposed Al-Cu-Li alloys [9-11]. Pit initiation, dissolution and sustained growth generate the surface geometry observed after exposure [12-13]. Nominally identical exposure can produce distinct pits, neighbouring depressions and elongated grooves, making instance definition important when comparing dimensions. Three-dimensional observations help distinguish their lateral extent from penetration into the material. Time-resolved X-ray microtomography has followed localized attack in a high-strength aluminium alloy [14]. Infinite-focus microscopy and confocal surface metrology provide optical routes to quantitative pit descriptors [15-16]. Sasidhar and colleagues recently investigated convolutional neural networks for automated analysis of nondestructively acquired three-dimensional corrosion-pit morphology [17]. These developments motivate a more specific question than whether machine learning can recognize corrosion. How do recovered instances and their boundaries affect the physical quantities reported from a measured surface?

Corrosion-image analysis has progressed from broad recognition towards localized segmentation and reliability assessment [18-19]. Recent work compares YOLO segmentation models on metal surfaces, encoder-decoder models for corroded steel structures, and localization of individual pitting features [20-23]. Other studies segment rust layers or combine optical measurements with learning-based analysis of corrosion susceptibility [24-26]. Their materials, image modalities and corrosion definitions differ, so pixel-level accuracy is not interchangeable with recovery of separate pits. Zhao and colleagues addressed localized corrosion segmentation in structural alloys through dedicated annotations and texture refinement [27]. Such boundary analysis acquires an additional physical role when calibrated confocal heights are available. A foreground map can merge adjacent defects, whereas instance masks preserve the spatial support and identity of each measured pit. Each accepted mask selects the pixels used to calculate its area and depth. Consequently, boundary quality and instance recovery require separate evaluation, particularly when shallow rims surround well-resolved pit bottoms.

Instance-segmentation architectures provide spatial supports for individual defects rather than one undifferentiated foreground region [28-29]. Transformer detection and promptable segmentation broaden the available localization interfaces [30-31], while vision-language models can generate additional object proposals [32-34]. Nash, Zheng and Birbilis showed that uncertainty estimates help interpret corrosion predictions beyond a single accuracy score [35]. These advances also expose important differences between recognition outputs and physical measurements. A bounding box includes surrounding surface, normalized intensity does not express depth, and a detection score does not define a pit boundary. Combining a segmenter with a proposal generator therefore requires explicit roles for their different spatial outputs. Additional proposals may recover missed pits but also admit grooves or texture, changing the population subsequently measured. A measurement-oriented interface must connect accepted instances to the original height grid while distinguishing population recovery from geometric agreement within matched pits.

Here we develop PitRPM, which connects learned recognition to geometric computation on the original confocal height field. A three-channel representation encodes surrounding relief, local depression and boundary structure. A high-resolution segmenter supplies instance masks, while a task-adapted Florence-2 branch supplies additional boxes through an asymmetric fusion interface. Spatial overlap and local depth govern candidate admission, and a common measurement operator converts accepted masks into physical descriptors. Using reviewed annotations from 120 fields of 30 corroded 2195-T8 specimens, we evaluate instance recovery, geometric agreement and the detection balance across model configurations.

Results

Connecting recognition to physical measurement

PitRPM separates the representations used for recognizing and measuring a pit (Fig. 1). The learned branches receive normalized height, nonnegative depth residual and gradient magnitude, whereas geometric computation retains the original heights and lateral spacing. YOLO11x-seg supplies instance masks, and Florence-2 supplies additional boxes whose support is recovered from the local depth residual. Their interface is asymmetric rather than an average of competing predictions. Enabling the proposal branch preserves the delineation of retained segmenter masks. An additional box identifies a region for residual-based mask construction and is admitted only after overlap and depth screening. It does not replace the nearest segmentation boundary or supply a depth value directly. All accepted masks then enter the same measurement operator on the original spatial grid. Figure 1 distinguishes boxes from masks because coordinate sequences localize candidate regions, whereas pixel supports determine which measured surface values enter each descriptor. This organization connects complementary recognition outputs without making physical scale dependent on image normalization or generated box coordinates.

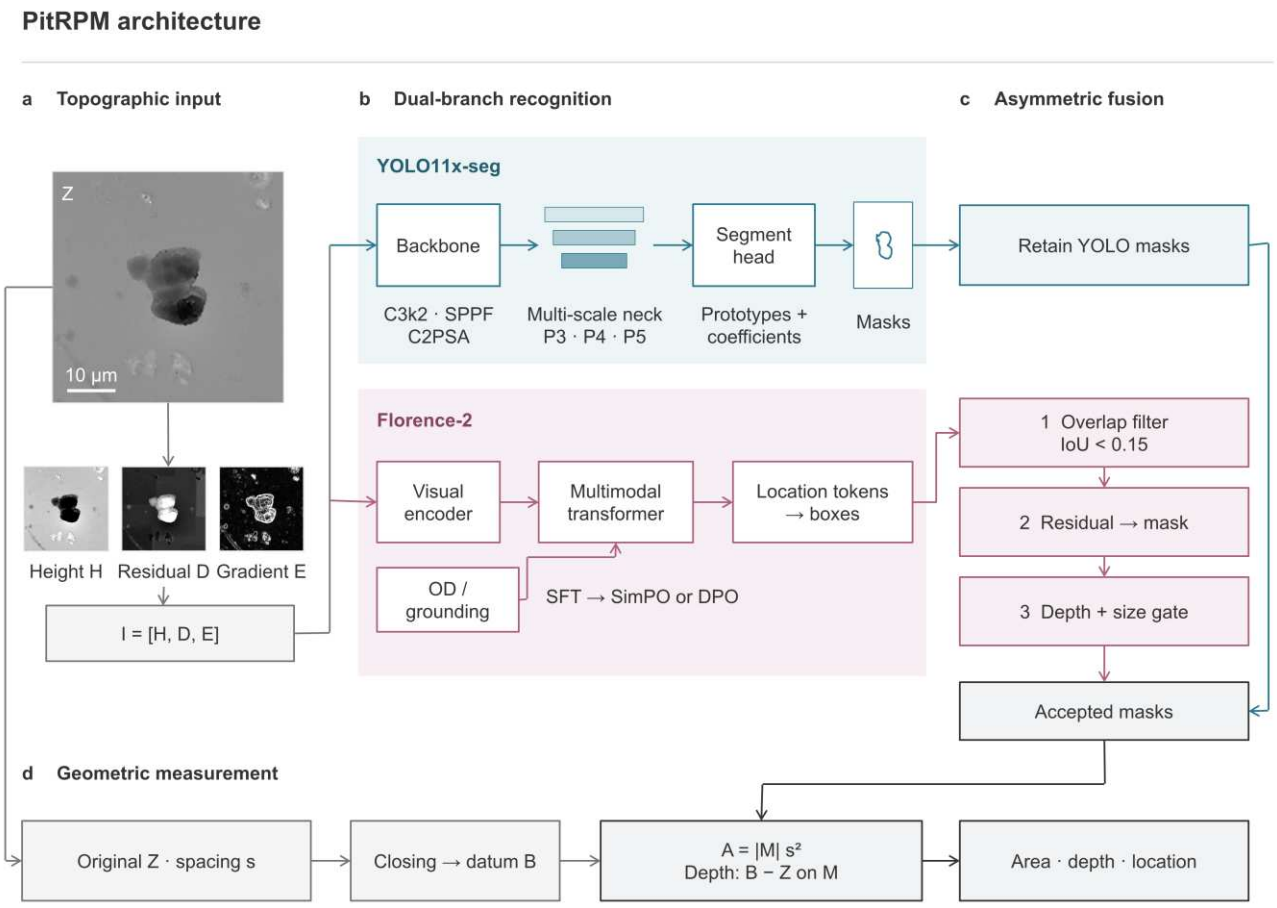


Fig. 1. Recognition and measurement pathways in PitRPM. The segmenter supplies masks and Florence-2 supplies optional boxes. Overlap and depth screening govern additional candidates. All accepted masks are measured using the original height field and local datum. SimPO and DPO are alternative post-SFT configurations.

The input channels describe complementary aspects of the same registered surface (Fig. 2). Height preserves surrounding relief, the residual emphasizes depressions, and the gradient emphasizes rims and grooves. Corresponding pixels refer to the same location, so these channels provide surface context, local depth contrast and boundary structure without combining separate acquisitions. Their roles follow the representation design rather than independently measured channel contributions. A depression is encoded relative to both its surroundings and the local datum, while its boundary is represented through spatial height changes. The retained physical field remains available for measurement after recognition, distinguishing encoded contrast from depth in micrometres. Evaluation used 96 training fields containing 889 pits and 24 validation fields containing 249 pits.

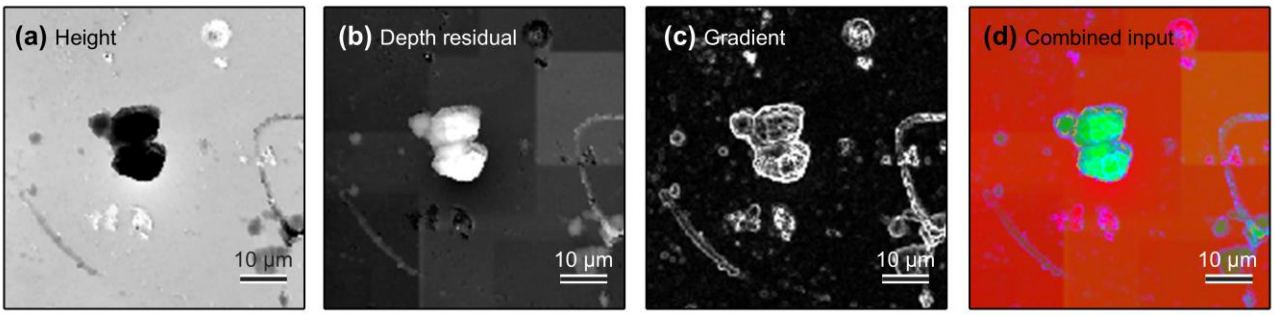


Fig. 2. Three-channel representation of one confocal field. (a-c) Normalized height, depth residual and gradient magnitude. (d) Their combined input, with reviewed reference contours. All panels use the same crop from field 002_2; the input channels are dimensionless. Scale bars, 10 µm.

Instance recovery and matched-pit geometric agreement

The segmenter-only configuration achieved the highest F1 among the evaluated configurations (Table 1). At confidence 0.25, YOLO11x-seg returned 211 true positives, 51 false positives and 38 false negatives, corresponding to precision 0.805, recall 0.847 and F1 0.826. Its matched-mask mean intersection-over-union (mIoU) was 0.758. Otsu thresholding was the strongest classical comparator by F1, reaching 0.253. All methods used the same local-datum definition for geometric measurement; the comparison therefore concerns recognition and mask support under that shared operator.

Table 1. Detection and matched pit measurement results. Validation includes 24 fields and 249 reference pits. RE, mean relative error expressed as a fraction; mIoU, matched mask mean intersection-over-union; CC, connected components; ASF, alternating sequential filtering. All methods share the same geometric measurement operator.

Method	Precision	Recall	F1	mIoU	Area RE	Max-Depth RE	Mean-Depth RE
Depth-thr 0.30 + CC	0.028	0.357	0.052	0.601	0.858	0.0080	0.187
Otsu on D + morphology	0.159	0.622	0.253	0.706	0.513	0.0149	0.090
Top-hat/ASF residual	0.027	0.257	0.049	0.596	0.840	**0.0050**	0.177
YOLO11l-seg (conf 0.25)	0.770	0.791	0.780	0.751	**0.288**	0.0118	**0.050**
YOLO11x-seg (conf 0.25)	**0.805**	0.847	**0.826**	0.758	0.318	0.0078	0.052
+ Florence-2 (SFT, gated)	0.670	0.863	0.754	0.757	0.323	0.0081	0.052
+ Florence-2 (SimPO, gated)	0.678	0.855	0.757	0.759	0.317	0.0082	0.051
+ Florence-2 (DPO, gated)	0.671	0.851	0.750	**0.761**	0.310	0.0082	**0.050**
+ Florence + OWL + DINO (gated)	0.595	0.871	0.707	0.757	0.316	0.0092	0.051
+ Florence + OWL + DINO (no gate)	0.524	**0.888**	0.659	0.753	0.323	0.0076	**0.050**

The learned configurations provided different precision-recall balances, with YOLO11x-seg exceeding YOLO11l-seg in F1 under the reported training settings (Fig. 3). Classical methods sometimes identified matched depressions with close depth agreement, but their lower F1 scores reflected poorer recovery of the instance population. Small matched-depth error therefore did not imply an accurate field-level inventory. The segmentation models recovered a more balanced population while retaining boundary-dependent differences in projected area. Matched-mask overlap adds a separate measure of shared spatial support within accepted correspondences.

Precision and recall describe which pits are recovered, mIoU describes their overlap, and descriptor errors describe physical quantities calculated from corresponding masks. Table 1 reports these levels together rather than inferring overall characterization quality from one favourable metric.

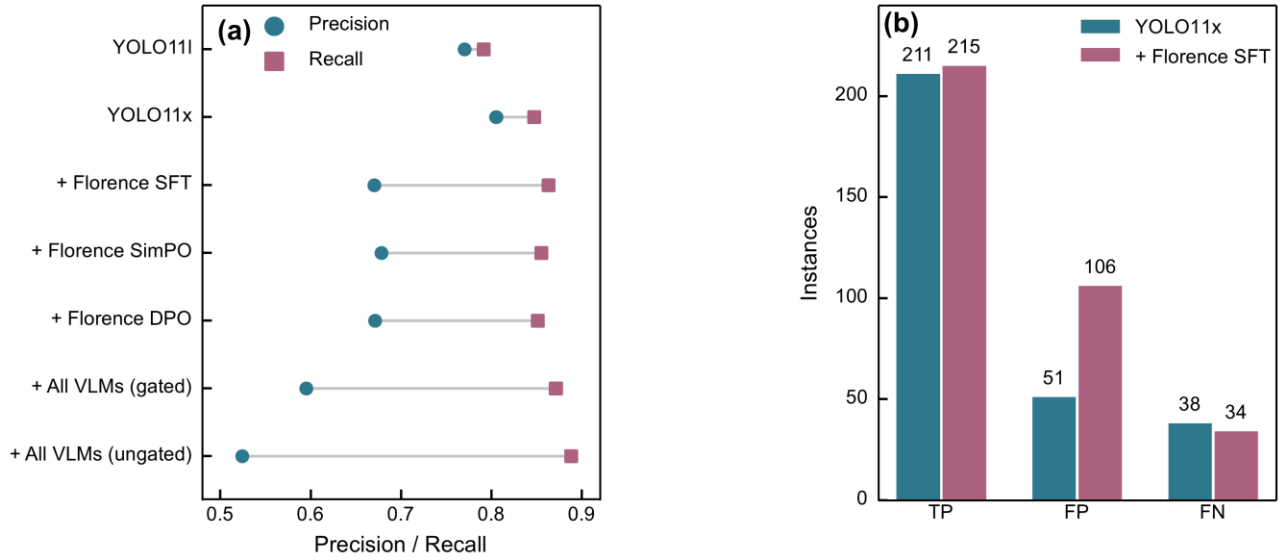


Fig. 3. Instance-detection performance. (a) Precision and recall across learned configurations. (b) True-positive, false-positive and false-negative counts for segmentation alone and Florence-2 supervised fine-tuning (SFT) fusion. Results refer to 24 validation fields containing 249 reference pits.

The confidence sweep placed the selected operating point near the highest validation F1 score (Fig. 4). Raising or lowering this threshold changes which predicted instances enter the output population and therefore changes precision and recall. Confidence 0.25 specifies the acceptance rule used for the subsequent geometric comparison, not a physical depth threshold. It does not alter the height assigned to any measured surface location. Segmentation alone consequently provided the balanced detection configuration used for comparing geometric descriptors.

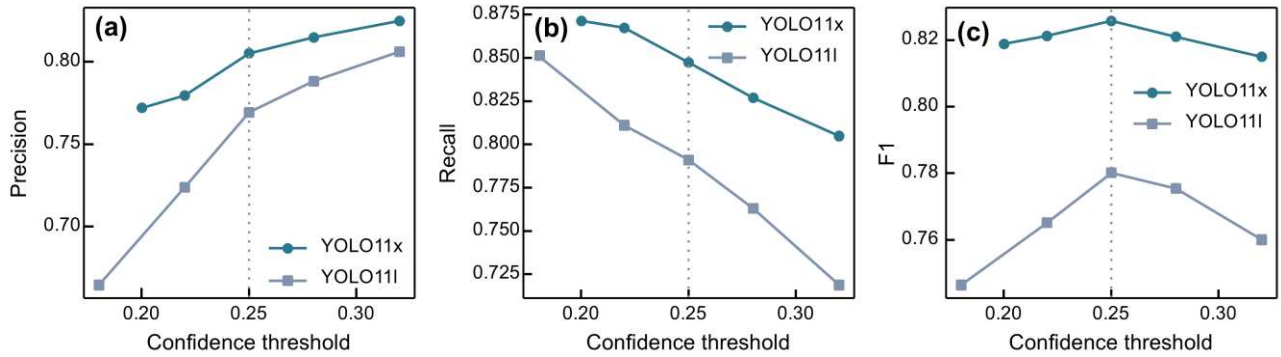


Fig. 4. Confidence-threshold behaviour. Precision, recall and F1 for the two segmenters, redrawn from archived vector traces. The vertical line marks confidence 0.25. The recorded sweeps begin at 0.20 for YOLO11x and 0.18 for YOLO11l.

Boundary placement affects area more than depth

For matched YOLO11x-seg instances, mean relative error was 31.8% for projected area, 0.8% for maximum depth and 5.2% for mean depth. These descriptors weight the same mask differently. Projected area counts every included pixel equally, irrespective of its position near the rim or

bottom. Maximum depth selects the deepest finite residual, while mean depth averages residual values over the valid support. A rim displacement can therefore change area appreciably without changing maximum depth when the pit bottom remains included. Adding or omitting shallow peripheral pixels also changes the mean depth. The reported pattern is consistent with retaining the deepest region more faithfully than the complete footprint. Predicted and reviewed masks sample the same acquired height field and local datum. Their descriptor differences consequently quantify mask-supported agreement rather than independent instrumental accuracy. A close maximum-depth match indicates similar deepest residuals within the retained supports. The area comparison asks a broader spatial question because it incorporates the complete footprint. Reporting both preserves the distinction between locating the bottom neighbourhood and recovering the surrounding boundary.

Figure 5 connects the local datum and instance mask to individual and population-level descriptors. The reviewed example has projected area 151.125 μm², maximum depth 5.44 μm and mean depth 3.01 μm. Its spatial panels make the selected footprint and cross-sectional depths inspectable at one location. The population panel instead contains all 262 segmenter predictions, comprising 211 matched and 51 unmatched instances. It is an output-population view, not a distribution restricted to confirmed correspondences. This distinction links the geometry plot to the detection counts without letting one representative pit imply the recognition status of every prediction.

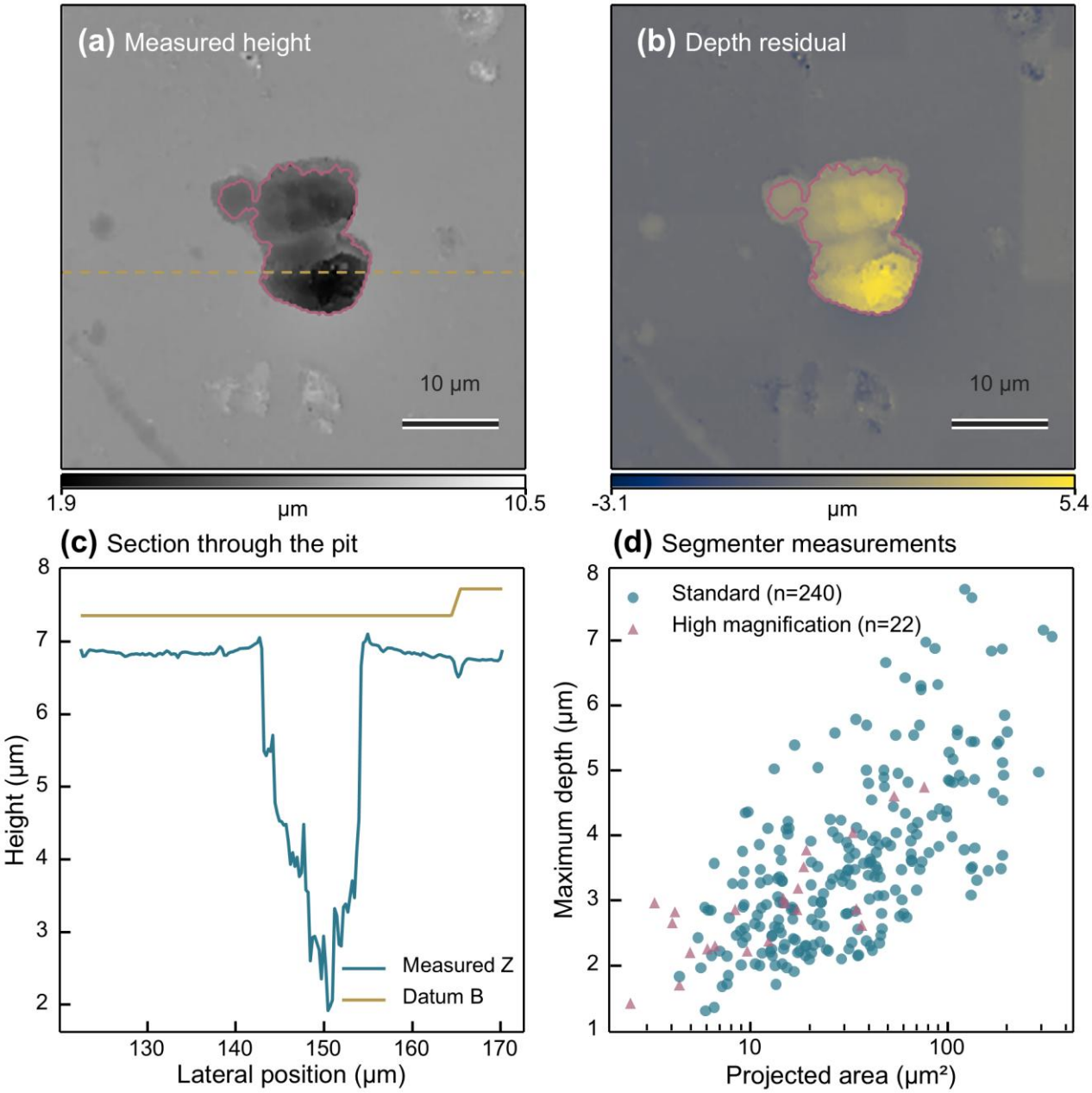


Fig. 5. From a reviewed pit to geometric descriptors. (a,b) Height and signed depth residual with the reference mask. (c) Surface and datum along the dashed section. (d) Area and maximum depth for all 262 segmenter predictions, including 51 unmatched predictions. Panels a-c use field 002_2, instance 8. Scale bars, 10 µm.

Representative interior, high-magnification and field-margin morphologies illustrate the reviewed instance definition (Fig. 6). Paired height and contour panels retain the same measured surface beneath each annotation. Interior regions show neighbouring features, the high-magnification crop exposes boundary detail, and field-margin examples show the acquired support near an image edge. Calibrated scale bars distinguish physical feature dimensions from display enlargement. These examples document the morphology addressed by the annotations rather than additional detection experiments.

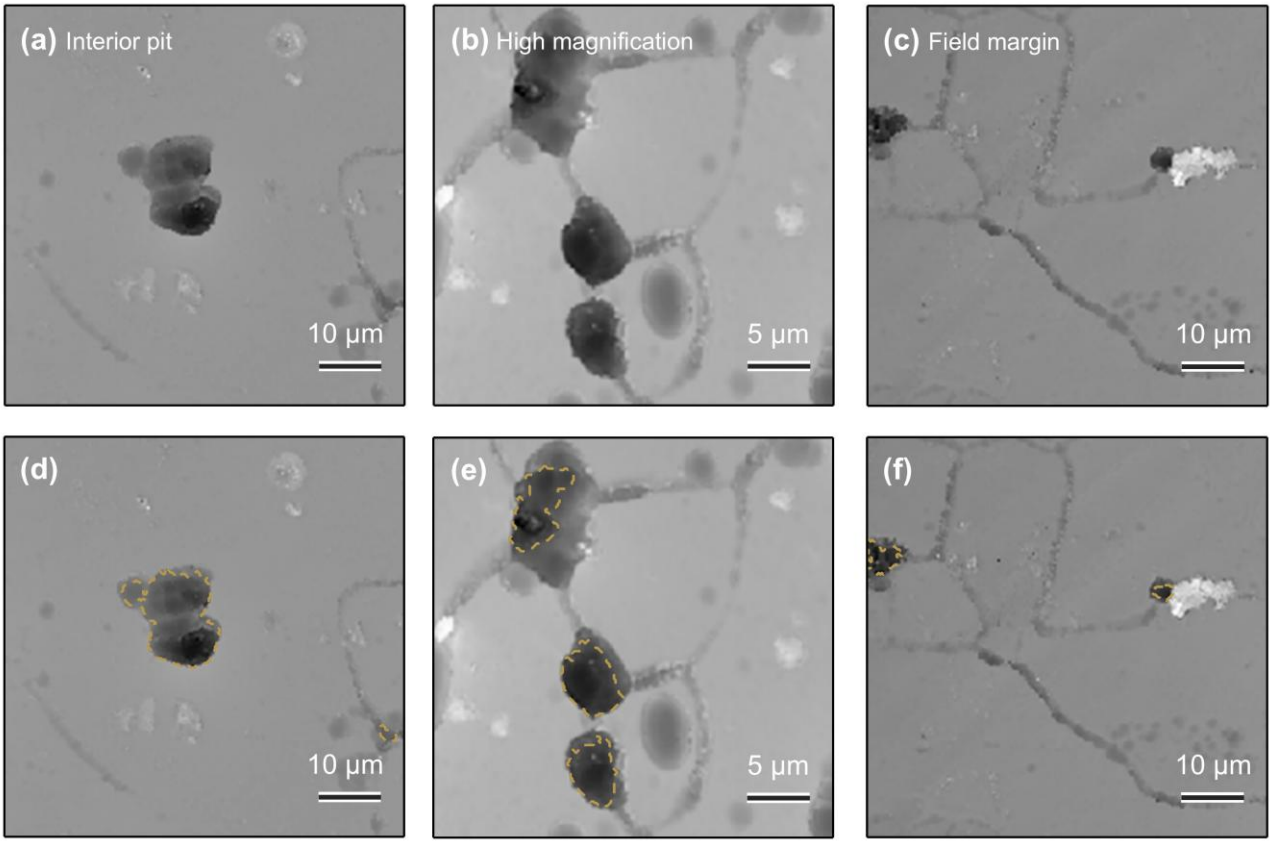


Fig. 6. Reviewed pit morphologies. Top row, measured height; bottom row, identical crops with reviewed instance contours. Columns show interior, high-magnification and field-margin regions. Scale bars, 10, 5 and 10 µm, respectively.

Candidate fusion adjusts the precision-recall balance

Fusion with Florence-2 after supervised fine-tuning (SFT) raised recall to 0.863, with precision 0.670 and F1 0.754. Relative to segmentation alone, fusion added four true positives and 55 false positives. Spatial examples distinguish retained segmenter masks from residual-derived additions on identical measured crops (Fig. 7). These routes produce the same final mask data structure but use different boundary-generation processes. The retained masks originate from supervised segmentation, whereas added masks originate from box proposals followed by residual-based support construction. Their colour distinction makes the asymmetric interface inspectable without redrawing existing segmentation boundaries. Local depression supplies admission evidence, although grooves and non-pit texture can also satisfy depth criteria. The tabled results quantify the resulting recognition balance.

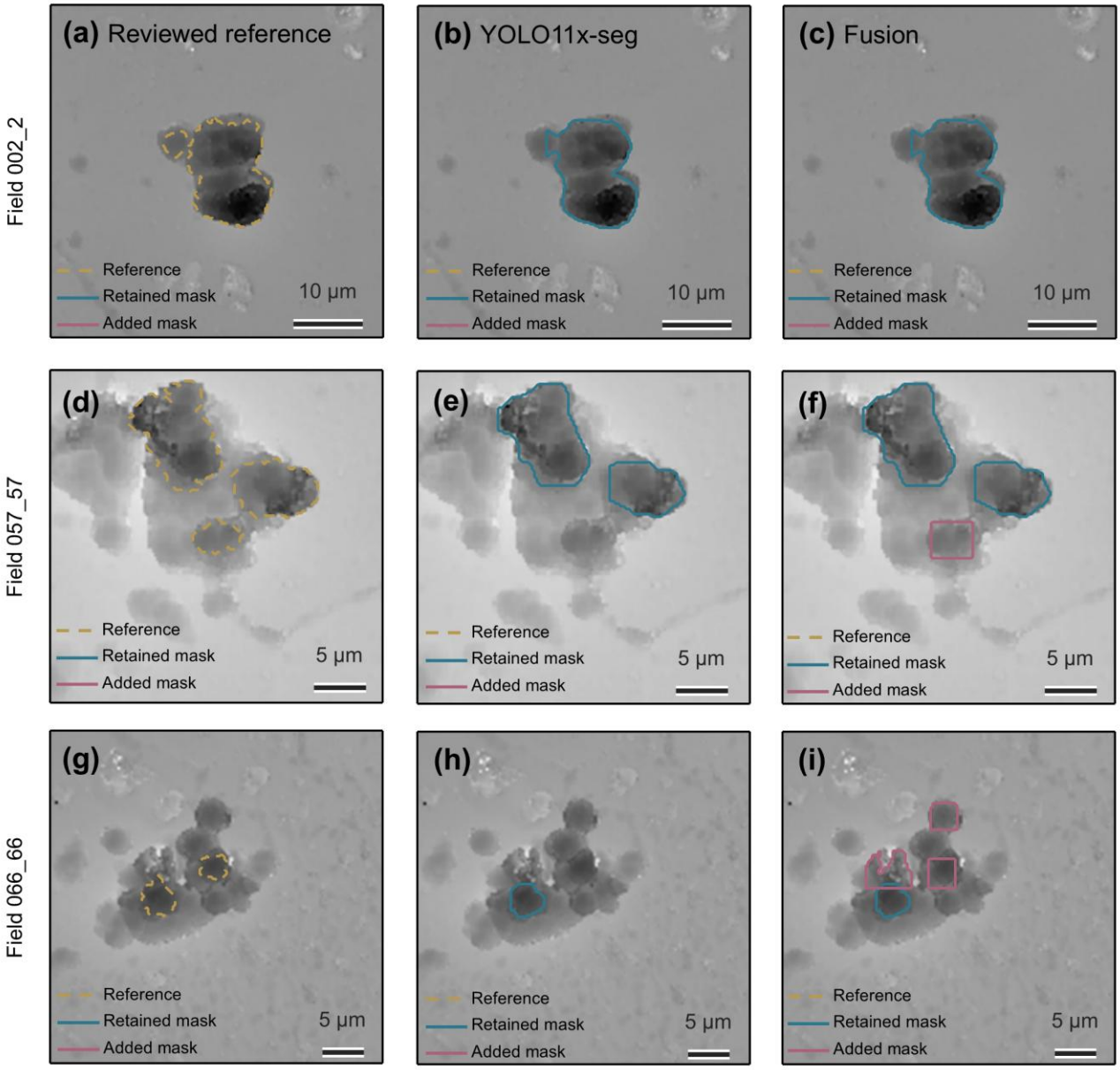


Fig. 7. Spatial examples of retained and added masks. Columns compare reviewed contours, segmenter output and fusion on identical height crops. Gold dashed contours denote reference instances; teal and berry denote retained and added masks. Rows show fields 002_2, 057_57 and 066_66. Scale bars, 10, 5 and 5 µm, respectively.

Simple preference optimization (SimPO) reduced false positives from 106 to 101 relative to SFT, while recall decreased from 0.863 to 0.855. F1 increased from 0.754 to 0.757, whereas direct preference optimization (DPO) yielded F1 0.750 (Table 2). These alternative post-SFT configurations modify the proposal sequences entering fusion, not the segmentation targets or geometric definitions. SFT establishes the localized detection format, and preference objectives favour selected sequences through task-derived comparisons. SimPO achieved the highest fused F1 among these alternatives, while DPO produced the highest matched-mask mIoU in Table 1. The rankings concern different output properties and do not establish interchangeable notions of improvement.

Table 2. Florence 2 adaptation and fusion results. SFT, supervised fine-tuning; SimPO, simple preference optimization; DPO, direct preference optimization; FP, false positives; OD, object detection. SimPO and DPO are alternative post-SFT configurations. Dashes denote an unreported numerical result.

Configuration	Precision	Recall	F1	FP ↓	Note
Zero-shot Florence-2	-	-	-	-	Domain-shift baseline; emits near-full-frame boxes
SFT only	0.670	0.863	0.754	106	Pulls output to localized OD format
SFT + SimPO (reference-free)	**0.678**	0.855	**0.757**	**101**	Best fused validation F1
SFT + DPO (with reference)	0.671	0.851	0.750	104	Classic control

For the larger Florence-OWL-DINO proposal set, geometric gating increased precision from 0.524 to 0.595 while reducing recall from 0.888 to 0.871. The gate selects candidates through spatial overlap and local depression rather than adding another learned segmentation network. Figure 8 shows how thresholded residuals define connected support inside a proposed box before final admission. The accepted support subsequently enters the common geometric measurement operator. Both preference adaptation and geometric screening therefore change which candidates are retained. None of the evaluated fusion configurations surpassed the balanced F1 of segmentation alone.

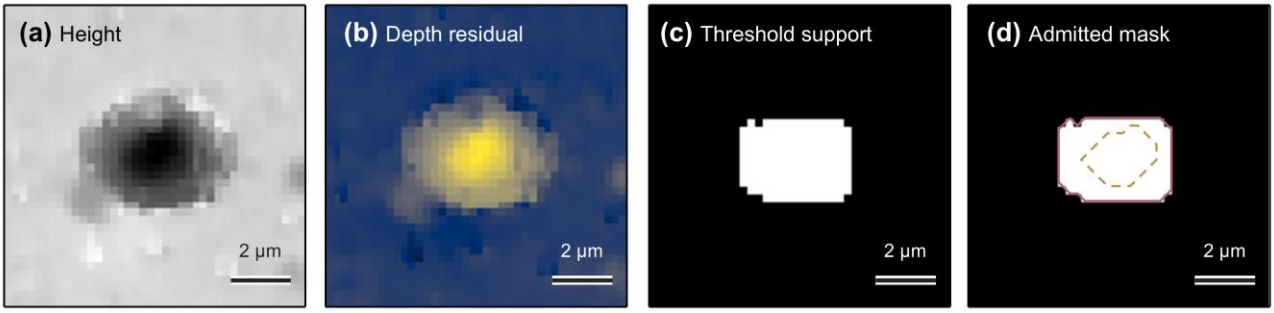


Fig. 8. Residual-based proposal mask construction. Panels show measured height, depth residual, threshold support and final admitted mask in field 049_49. Dashed and solid contours denote the reviewed reference and admitted mask. The example uses a saved proposal from the same local run as main Fig. 7. Scale bars, 2 µm.

The resolution-tier summaries describe sampling and geometric agreement together (Table 3). The high-magnification validation subset contains four fields and 21 reference pits and shows greater area error than the standard subset. A boundary pixel represents different physical areas at the two sampling intervals, but descriptor errors also depend on the recovered footprints and residual distributions. The reported differences therefore characterize these subsets rather than isolating magnification as the cause of recognition performance. Tier-specific summaries are retained separately from the aggregate comparison in Table 1. Maximum-depth errors displayed as 0.0000 denote agreement at the reported numerical precision.

Table 3. Resolution tier comparison. GT, reference pit count; RE, mean relative error expressed as a fraction. Tier-specific counts do not reconstruct the aggregate counts in Table 1 and are presented as separate summaries. A value of 0.0000 denotes agreement at the displayed precision.

Tier	Images	GT	Precision	Recall	F1	mIoU	Area RE	Max-Depth RE	Mean-Depth RE
Standard (~0.25 µm/px)	20	228	0.747	0.855	0.798	0.763	0.306	0.0090	0.051
High-magnification (~0.125 µm/px)	4	21	0.708	0.810	0.756	0.743	0.354	0.0000	0.036

Discussion

PitRPM connects recognition, candidate admission and geometric computation through a shared spatial mask. The model-design contribution lies in organizing these decisions around acquired topography. YOLO11x-seg supplies instance boundaries, Florence-2 supplies structured proposals, and their asymmetric interface preserves masks while admitting selected additional candidates. The three-channel representation presents surface context, depression contrast and boundary structure to recognition without replacing the physical height field. The common operator then associates each accepted support with area and depth in calibrated units. These components answer successive questions about the same field: where a pit may be, which support is accepted, and what that support measures. Their separation is important because a recognition output becomes a quantitative corrosion record only after its spatial support and physical measurement rule are defined. The resulting architecture ends with traceable instance descriptors rather than a class label or box alone.

This design complements corrosion-imaging approaches that address boundary representation, prediction reliability and physical morphology at different levels. Zhao et al. used dedicated annotations and texture refinement to improve localized corrosion segmentation [27]. Nash et al. examined Bayesian prediction variability and pixel-level uncertainty [35], while Sasidhar et al. analysed three-dimensional pit morphology through learned sharpness representations [17]. PitRPM instead connects field-level instance recovery to direct area and depth computation from each accepted support. This comparison is particularly informative at the boundary between a pit and its surroundings. Texture refinement addresses continuity and contrast, whereas a confocal contour additionally selects numerical heights relative to a local datum. A displacement across a shallow rim changes the footprint differently from exclusion of a deep interior region. Segmentation quality thus becomes descriptor-specific rather than implying that all physical outputs follow the same overlap ranking. Task-specific detection and geometric metrics keep these questions separate, consistent with image-analysis validation guidance [36-37].

This descriptor-specific view also clarifies the role of confidence-aware detection. Bayesian variability describes uncertainty in model output, whereas a physical residual describes separation between a measured surface and its datum [35]. These quantities have different meanings and units, even when both inform interpretation. A deep candidate is not automatically a certain pit prediction, and a confident prediction is not automatically a deep feature. Likewise, the operating-threshold

sweep describes the detection balance associated with an acceptance rule, not uncertainty in measured depth. The geometric gate therefore contributes topographic evidence without converting depth into prediction probability. This distinction helps interpret hybrid outputs without conflating model confidence, physical depression and agreement with the reviewed pit definition.

The measured descriptors characterize the geometric outcome of localized material loss after a defined exposure. Chemical imaging has resolved corrosion-associated changes around intermetallic particles and differences between particle and matrix regions in Al-Cu-Li alloys [7]. Particle-associated evolution and processing-dependent corrosion connect local surface morphology to material heterogeneity [8-9,11]. Studies of precipitate-associated attack and pit dissolution further show the importance of microstructure and exposure conditions [12-13,38]. Figure 9a links the measured surface to the footprint selected for geometric characterization without assigning a particular relief feature to a specific phase reaction. Alloy state and exposure describe how the surface was produced; area and depth describe the accessible relief remaining after cleaning. The common nominal concentration, temperature and duration give the measured population a defined experimental context. Morphological differences are interpreted within that context rather than converting one endpoint into a growth history. This ordering connects the quantitative survey to corrosion science while distinguishing topographic evidence from chemical identification of the constituents involved in dissolution.

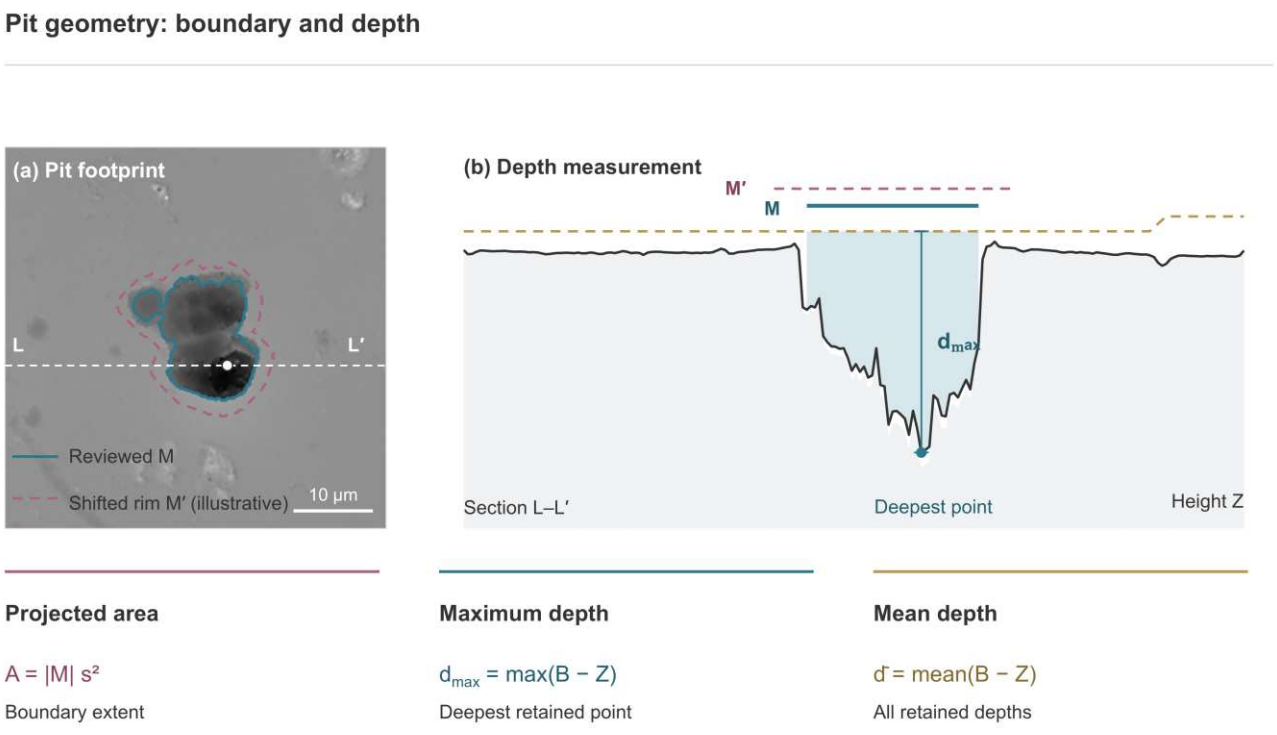


Fig. 9. Geometric interpretation of pit measurements. (a) Reviewed footprint and illustrative rim displacement on a measured height crop. (b) Surface profile and depth relative to the local datum. Equations summarize projected area, maximum depth and mean depth. The shifted rim is illustrative. Scale bar, 10 µm.

The different geometric errors reflect how each operator uses its mask support (Fig. 9a,b). Area integrates the entire footprint, making it sensitive to displacement along an extended or irregular rim. Pixels added on one side and omitted on another can partly cancel in the net pixel count even

when the contours disagree. Conversely, a coherent rim displacement can alter many pixels while leaving the deepest region untouched. Mask overlap and area error are therefore related but non-identical measures. Overlap describes shared support, while area error describes the net footprint difference multiplied by squared lateral spacing. Their joint reporting is useful for irregular pits with substantial boundary length relative to footprint area. The common acquired field and datum ensure that this comparison concerns delineation, complementing the calibrated topographic measurements used in optical pit characterization [14-16].

Maximum depth has an extremum-based sensitivity rather than the cumulative sensitivity of area. If predicted and reviewed masks retain the same deepest location, their maximum depths agree under a common datum despite differences elsewhere. Excluding that location substitutes the next deepest included residual and can change the result abruptly. Mean depth responds to both the selected footprint and its residual distribution. Adding shallow peripheral pixels increases area while lowering mean depth, even if maximum depth remains unchanged. Removing those pixels produces the opposite changes, so the three descriptors need not share an error scale. Datum consistency matters separately because an offset in the reference surface changes residual depths even when a mask remains fixed. The small matched maximum-depth discrepancy therefore indicates agreement in retained bottom regions on the same measured field, not independent verification of instrumental accuracy.

Beyond individual geometry, field-level characterization depends on population selection, which matched geometric errors alone do not describe. A missed pit contributes no descriptor to the predicted record, regardless of depth agreement among the pits that were recovered. An unmatched prediction contributes a descriptor whose membership is unsupported by the accepted correspondence. Detection metrics quantify these population differences, while geometric errors describe corresponding instances after matching. Figure 5d retains both matched and unmatched predictions so that its distribution can be interpreted with the recognition results. Instance identity further preserves information that a pooled foreground region would combine. Separate neighbouring masks produce separate measurement records; merging them changes the count and the footprint assigned to an object. This distinction is consistent with the relevance of neighbouring defects to pit-to-crack behaviour [6]. Retaining masks, coordinates and descriptors together consequently connects model comparisons in corrosion segmentation to inspectable physical outputs [20-21,37].

These population differences explain why fusion defines distinct operating modes rather than a uniform accuracy gain. Segmentation alone provides the highest balanced F1, while proposal fusion recovers additional reference pits with a larger false-positive population. The former suits a balanced automated inventory; the latter offers broader candidate recovery with an explicit increase in review burden. The mask-preserving interface keeps the initial segmentation set identifiable while adding candidates through a separate route. Preference adaptation and

geometric screening act at successive stages of that route [39-40]. Adaptation changes which detection sequences are generated, whereas the gate examines their spatial overlap and measured depression. A preferred sequence can still yield a rejected box, and an admitted depression can still be non-pit texture under the annotation definition. Their contributions are therefore not interchangeable. The common operator preserves area and depth definitions across these operating choices, making changes in the measured population attributable to recognition and admission rules.

Traceability also requires branch-specific scores and physical sampling to retain their distinct meanings. Equal confidence values from a segmenter and a proposal generator do not establish equal probabilities of a correct pit prediction. PitRPM uses branch-specific acceptance and spatial screening rather than averaging scores onto an assumed common calibrated scale. This keeps the processing route of an accepted instance interpretable. Learning-assisted optical monitoring, hyperspectral characterization and pitting detection similarly rely on information specific to their acquisition modalities [23,25-26]. For confocal fields, lateral spacing defines the area represented by each pixel, while the retained height residual supplies depth. The resolution-tier summaries therefore reflect sampled morphology, recovered support and measurement definitions together, rather than magnification alone. Calibrated scale bars and recorded spacing are integral to reading both the spatial examples and the resulting descriptors. Changing lateral sampling alters the physical area represented by a boundary pixel without directly defining the residual at the deepest point.

Together, recognition and geometric measurement provide pit-resolved descriptions relevant to structural degradation, where localized defects concentrate material loss within small surface footprints. Pit-to-crack studies connect mechanical response to pit shape, neighbouring defects, material state and loading [4-6]. Time-resolved tomography and in situ observations add the temporal dimension of corrosion evolution [14,41]. The descriptors summarized in Fig. 9 provide geometric inputs considered alongside mechanical and microstructural evidence. Two pits can share projected area while differing in maximum or mean depth, so no single descriptor captures their full geometrical distinction. Area records lateral extent, maximum depth records deepest penetration relative to the datum, and mean depth describes the average residual over the selected support. Interpreting these inputs in a structural application remains connected to its loading conditions and material evidence. Automated topographic characterization supplies these quantities consistently for individual defects after a defined exposure. For the evaluated 2195-T8 surfaces, PitRPM makes this route explicit by associating each accepted instance with its support, source height field and descriptors. It supports inspection of both the population selected for measurement and the geometry assigned to each member.

Methods

Specimens and confocal data processing

Thirty 2195-T8 aluminium-lithium alloy specimens, each 20 × 20 mm, were ground and polished before exposure to aqueous 5 wt% HNO_3. Exposure lasted 2 h at a constant 25 °C, followed by ultrasonic cleaning in anhydrous ethanol and blow-drying. An Olympus LEXT OLS4100 confocal laser scanning microscope recorded one field from each of four regions per specimen, producing 120 height fields. Consecutive groups identify specimens: fields 1-4 belong to specimen 1, fields 5-8 to specimen 2, and so on. Each field contains a 1024 × 1024 height grid. The standard tier comprises 100 fields at approximately 0.25 μm lateral sampling; the high-magnification tier comprises 20 fields at approximately 0.125 μm. Pits were manually annotated as separate instances and manually reviewed, excluding grain-boundary grooves, raised products and very shallow texture. The resulting 1,138 masks define the segmentation references and support reference geometric measurements on the acquired heights. A seed-42 split, stratified by resolution tier, assigned 96 fields containing 889 pits to training and 24 fields containing 249 pits to validation. Partitioning was by field rather than specimen, allowing fields from one specimen in both subsets. Validation informed model development and threshold characterization, and reported performance refers to this field-level evaluation [42].

Let Z denote the acquired height field, s the lateral pixel spacing and B the local surface datum. Morphological closing [43] estimates B from a fourfold-subsampled height field using a square footprint with a nominal 20 μm radius. With s in μm per pixel, the subsampled footprint radius is max{1, floor[round(20/s)/4]}. The datum is interpolated to the original grid before cropping and is shared by predicted and reviewed masks within each field. Its footprint therefore follows the same nominal physical rule at both resolution tiers. Datum construction precedes recognition and is not adjusted according to detection confidence. Using one datum also prevents descriptor differences from arising through independently fitted reference surfaces for predicted and reviewed masks. Invalid heights are replaced by the median finite height only when constructing the recognition input, giving $\tilde{Z}$. The input channels encode normalized height, nonnegative depth residual and Sobel gradient magnitude on the registered grid. Let Q_k denote percentile k and G the magnitude of the two 3 × 3 Sobel responses. The channels are defined as

$$H = \mathrm{clip}\left(\frac{\tilde{Z} - Q_1(\tilde{Z})}{Q_{99}(\tilde{Z}) - Q_1(\tilde{Z})}, 0{,}1\right),$$

$$D = \mathrm{clip}\left(\frac{\max(B - \tilde{Z}, 0)}{Q_{99.5}[\max(B - \tilde{Z}, 0)]}, 0{,}1\right), \qquad E = \mathrm{clip}\left(\frac{G}{Q_{99}(G)}, 0{,}1\right).$$

Denominators are floored at 10^{-12}, and the ordered channels H, D and E are stored as 8-bit values. Percentile normalization sets recognition contrast, with clipping bounding each channel while leaving the separately retained physical field unchanged. Height uses its central percentile range,

while the residual and gradient use their respective upper-percentile scales. The nonnegative input residual suppresses values above the datum, and channel order remains fixed during training and inference. For a mask M, projected area is $A = |M|s^2$, where |M| is its pixel count. This quantity is the top-down footprint, not the developed three-dimensional surface area. Maximum and mean depth are the maximum and average finite values of B−Z within M. The measurement residual is signed, unlike the nonnegative residual used for input encoding. Missing residuals are excluded rather than assigned the replacement value used for model input. Reference descriptors use reviewed masks on the same original heights, physical spacing and datum. Each predicted descriptor and its reference counterpart are calculated independently from their respective supports under this common operator. Geometric comparison therefore concerns physical residuals and footprint areas, not normalized intensities or box dimensions.

Model training and fusion inference

The segmentation branch uses YOLO11x-seg with 1024-pixel input and full-resolution mask output. Reviewed labels are converted into separate polygon targets for one pit class, preserving neighbouring instance identities. The branch returns boxes, confidences and masks; masks provide measurement support, while original heights remain outside the recognition input interface. No additional backbone block is introduced by fusion. Training used AdamW [44], cosine learning-rate scheduling, mixed precision and 120 epochs. The initial learning rate was 8×10^{-4}, weight decay 0.04 and warm-up five epochs. Mosaic probability was 0.3, with mosaic disabled during the final 20 epochs. Rotation was limited to 5°, translation to 0.05 and scale augmentation to 0.25; mixup was disabled. Confidence 0.25 was the reported operating point. YOLO11l-seg used the same field split as the comparative segmentation configuration.

Florence-2 [32] receives the same channels and generates structured box proposals, not measurement masks. Supervised fine-tuning maximizes the conditional likelihood of detection sequences constructed from reviewed boxes and quantized location tokens. These tokens specify candidate coordinates rather than descriptions of pit geometry. Training used 12 epochs and an initial learning rate of 5×10^{-6}. The visual encoder was initially frozen and subsequently unfrozen with the learning rate multiplied by 0.3; non-finite loss steps were skipped. At inference, object detection and phrase grounding with the prompt “corrosion pit dark depression hole pitting” supply candidate boxes. Within-branch deduplication uses box intersection-over-union (IoU) 0.4. Boxes enter spatial screening before residual support is constructed for measurement. Both learned branches use the same registered topographic input, keeping their distinct recognition routes connected to the same acquired surface.

Simple preference optimization (SimPO) [40] and direct preference optimization (DPO) [39] were evaluated as separate three-epoch post-SFT alternatives. Task-derived preferences combine box F1 and matched-box IoU with penalties for unmatched and near-full-frame boxes. Candidate sequences are compared within an image to select preferred and rejected outputs. DPO compares

their likelihood ratios against a frozen SFT policy, anchoring the comparison to that reference. SimPO uses length-normalized sequence likelihood and a target margin without a reference policy. They are alternative adaptations rather than consecutive stages of one final model. Their outputs remain box proposals processed by the same fusion rules. Figure 10 shows recorded optimization behaviour for the segmenter and proposal adaptations. Objective-specific loss scales describe their respective training objectives and are not combined into a common accuracy or geometric-error measure.

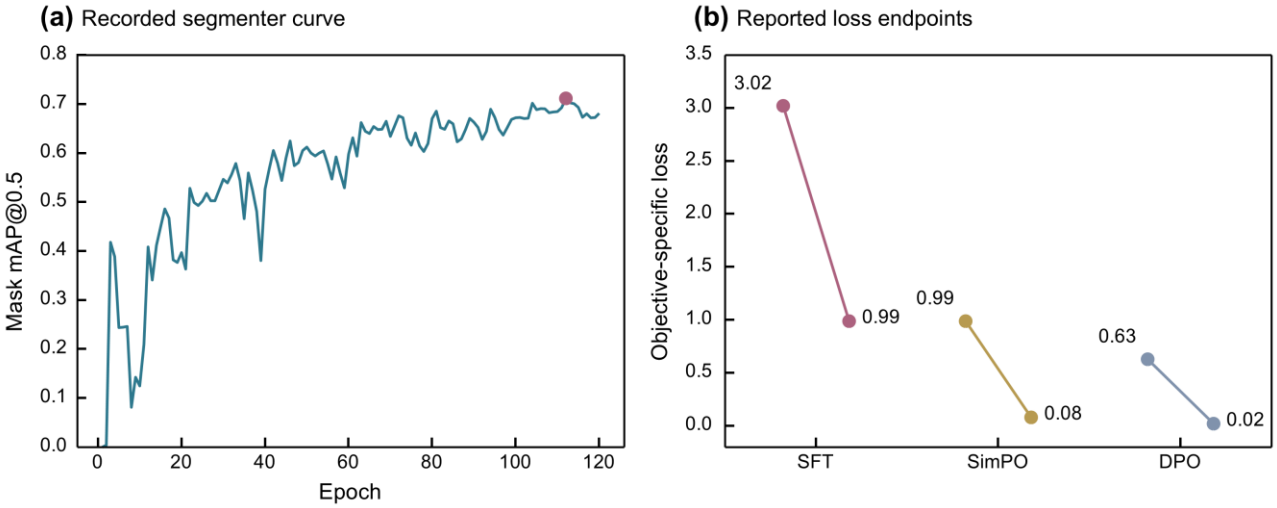


Fig. 10. Recorded optimization behaviour. (a) Segmenter mask mean average precision at IoU 0.5 (mAP@0.5), with the reported best checkpoint at epoch 112. (b) Loss endpoints for SFT and alternative preference objectives. Lines connect start and end values; the objective-specific losses use different scales.

Fusion initializes the retained set with segmenter boxes and masks. Additional proposals are processed sequentially, with Florence-2 preceding OWLv2 [33] and Grounding DINO [34] when included. A proposal passes overlap screening when its box IoU with every retained box is below 0.15. Earlier additional proposals participate in this comparison before geometric screening. The additional-proposal confidence floor is min(cY, 0.12), where cY is the segmenter threshold; scores remain branch-specific rather than commonly calibrated. An additional box must have area ≥16 pixels, with neither dimension exceeding one quarter of the corresponding image dimension. Its 85th-percentile finite depth residual must reach 0.7 μm. Within the box, pixels exceeding max(0.35 μm, $0.5Q_{70}$) define foreground, where Q_{70} is the 70th-percentile residual. The largest connected component is retained and holes are filled. Empty foreground triggers a retry at max(0.35 μm, Q_{50}), with the box region supplying fallback support if needed. Final admission requires at least 20 mask pixels, maximum depth ≥0.8 μm and mean depth ≥0.45 μm. Overlap screening thus uses rectangular support, whereas final admission uses a constructed mask and its measured residuals. Percentile screening requires a depressed portion of the box rather than requiring every included pixel to exceed one depth. Pixel thresholds specify support sizes, whereas micrometre thresholds refer to the physical residual.

Segmenter masks are associated with retained boxes at box IoU ≥0.3, using each mask once. Unmatched segmenter masks with confidence ≥0.30 are also retained. Classical comparators use

depth thresholding, Otsu thresholding [45], or top-hat/alternating sequential filtering with connected-component processing [43]. Their minimum component size is 50 pixels and minimum solidity is 0.25. Comparator parameters were selected within the validation workflow.

Evaluation and reproducibility

Predictions and reviewed masks are assigned by Hungarian matching to maximize mask IoU, accepting matches at IoU ≥0.3. Precision, recall and F1 are calculated from matched-instance counts; mIoU averages overlap among matched masks. Unassigned predictions remain false positives and unassigned reference instances remain false negatives, without imputed geometric errors. For area or depth q, relative error is $|\hat{q}-q|/q$, averaged over matched instances. The reference q is computed from the reviewed mask on the same height field and datum. Errors are calculated per corresponding instance before averaging, retaining each reference quantity as its denominator. Detection metrics therefore cover the complete predicted and reference populations, while geometric errors answer a conditional question about accepted correspondences [36-37]. Experiments used one RTX 4070 Ti SUPER with 16 GB memory.

Tables 1-3 report the original experiments. Figure 7 and Fig. 8 illustrate fusion operations using saved predictions from a separate run on 26 September 2026. These qualitative examples are distinct from the quantitative evaluation summarized in the tables. Figure 5a-c uses a reviewed mask; Fig. 5d uses the archived segmenter measurement records. Confidence and training curves were redrawn from archived vector traces. Microscopy panels retain registered crop coordinates and calibrated scale bars, and the conceptual schematic was drawn as vector artwork. Figure-source records are retained with the manuscript.

Code availability

The custom code for data preprocessing, model training, inference and evaluation, and the trained model weights corresponding to the reported experiments have not been deposited in a public repository. During peer review, the code and weights will be provided to the editors and reviewers upon request through the corresponding author. Following publication, they will remain available to researchers upon request from the corresponding author.

Data availability

The confocal height fields, manually annotated and reviewed pit labels, and source data underlying the figures and tables have not been deposited in a public repository. During peer review, these data will be provided to the editors and reviewers upon request through the corresponding author. Following publication, they will remain available to researchers upon request from the corresponding author.

References

1. Polayya, C., Rao, C. S. P. & Kumar, G. B. V. Improving the properties of aluminum-lithium composites in aerospace applications: review. *Eng. Res. Express* **6,** 022501 (2024). https://doi.org/10.1088/2631-8695/ad3a37
2. Frankel, G. S. Pitting Corrosion of Metals: A Review of the Critical Factors. *J. Electrochem. Soc.* **145,** 2186-2198 (1998). https://doi.org/10.1149/1.1838615
3. Pao, P. S., Feng, C. R. & Gill, S. J. Corrosion Fatigue Crack Initiation in Aluminum Alloys 7075 and 7050. *Corrosion* **56,** 1022-1031 (2000). https://doi.org/10.5006/1.3294379
4. Katona, R. M., Karasz, E. K. & Schaller, R. F. A Review of the Governing Factors in Pit-to-Crack Transitions of Metallic Structures. *Corrosion* **79,** 72-96 (2023). https://doi.org/10.5006/4179
5. Turnbull, A., Wright, L. & Crocker, L. New insight into the pit-to-crack transition from finite element analysis of the stress and strain distribution around a corrosion pit. *Corros. Sci.* **52,** 1492-1498 (2010). https://doi.org/10.1016/j.corsci.2009.12.004
6. Fatoba, O. & Akid, R. The influence of corrosion pit-pit spacing on the pit-to-crack transition and fatigue lifetime. *Fatigue Fract. Eng. Mater. Struct.* **45,** 2676-2692 (2022). https://doi.org/10.1111/ffe.13755
7. Li, M., Seyeux, A., Wiame, F., Marcus, P. & Światowska, J. Localized corrosion induced surface modifications of Al-Cu-Li alloy studied by ToF-SIMS 3D imaging. *npj Mater. Degrad.* **5,** 23 (2021). https://doi.org/10.1038/s41529-021-00170-9
8. Lv, Y., Meng, X., Dong, Z., Ma, Y. & Zhang, X. Localized corrosion evolution associated with intermetallic particles in Al-Cu-Li alloys. *Mater. Charact.* **204,** 113144 (2023). https://doi.org/10.1016/j.matchar.2023.113144
9. Deng, C., Wang, C., Wang, F., Song, B. & Zhang, H. Effect of microstructure evolution on corrosion behavior of 2195 Al-Li alloy friction stir welding joint. *Mater. Charact.* **184,** 111652 (2022). https://doi.org/10.1016/j.matchar.2021.111652
10. Wang, J. *et al.* Localized corrosion behavior of the Sc microalloyed Al-Cu-Li alloy. *Vacuum* **220,** 112832 (2024). https://doi.org/10.1016/j.vacuum.2023.112832
11. Magnan, M., Zanna, S., Seyeux, A., Wiame, F. & Światowska, J. Corrosion evolution of aluminium-copper and aluminium-copper-lithium alloys in chloride solution. *Appl. Surf. Sci.* **710,** 163922 (2025). https://doi.org/10.1016/j.apsusc.2025.163922
12. Li, T., Wu, J. & Frankel, G. S. Localized corrosion: Passive film breakdown vs. Pit growth stability, Part VI: Pit dissolution kinetics of different alloys and a model for pitting and repassivation potentials. *Corros. Sci.* **182,** 109277 (2021). https://doi.org/10.1016/j.corsci.2021.109277
13. Li, T., Wu, J., Guo, X., Panindre, A. M. & Frankel, G. S. Activation energy of metal dissolution in local pit environments. *Corros. Sci.* **193,** 109901 (2021). https://doi.org/10.1016/j.corsci.2021.109901
14. Sinclair, D., Niverty, S. & Chawla, N. Understanding Initiation and Evolution of Pitting Corrosion in 7075-T6 Aluminum Alloy by Four-Dimensional X-Ray Microtomography. *Corrosion* **79,** 4-16 (2023). https://doi.org/10.5006/4192
15. Chen, Y. & Ju, L.-K. Method for Fast Quantification of Pitting Using 3D Surface Parameters Generated with Infinite Focus Microscope. *Corrosion* **71,** 1184-1196 (2015). https://doi.org/10.5006/1556
16. Leach, R. (ed.) *Optical Measurement of Surface Topography* (Springer, 2011). https://doi.org/10.1007/978-3-642-12012-1
17. Sasidhar, K. N., Ahuja, R., Lukas, C. & Sridharan, K. Convolutional neural network for automated quantitative analysis of non-destructively acquired three-dimensional corrosion pit morphology data. *Scr. Mater.* **262,** 116660 (2025). https://doi.org/10.1016/j.scriptamat.2025.116660
18. Nash, W., Drummond, T. & Birbilis, N. A review of deep learning in the study of materials degradation. *npj Mater. Degrad.* **2,** 37 (2018). https://doi.org/10.1038/s41529-018-0058-x
19. Ma, S., Du, Y., Wang, S. & Su, Y. Application of machine learning in material corrosion research. *Corros. Rev.* **41,** 417-426 (2023). https://doi.org/10.1515/corrrev-2022-0089

20. Casas, E., Ramos, L., Romero, C. & Rivas-Echeverría, F. A comparative study of YOLOv5 and YOLOv8 for corrosion segmentation tasks in metal surfaces. *Array* **22,** 100351 (2024). https://doi.org/10.1016/j.array.2024.100351
21. Das, A., Dorafshan, S. & Kaabouch, N. Autonomous Image-Based Corrosion Detection in Steel Structures Using Deep Learning. *Sensors* **24,** 3630 (2024). https://doi.org/10.3390/s24113630
22. Krysko, N. V., Shchipakov, N. A., Kozlov, D. M. & Kusyy, A. G. Identification and localization of pitting corrosion on metallic surface using deep learning. *CIS Iron Steel Rev.* **27,** 96-102 (2024). https://doi.org/10.17580/cisisr.2024.01.15
23. Suzuki, T., Otaka, N., Shibano, K., Fujimoto, Y. & Hagiwara, T. Detection of Pitting Corrosion in Stainless-Steel Sheet Pile Walls Using Deep Learning. *Corros. Mater. Degrad.* **7,** 23 (2026). https://doi.org/10.3390/cmd7020023
24. Wang, B. *et al.* Advanced multi-image segmentation-based machine learning modeling strategy for corrosion prediction and rust layer performance evaluation of weathering steel. *Corros. Sci.* **237,** 112334 (2024). https://doi.org/10.1016/j.corsci.2024.112334
25. Makogon, A. *et al.* Machine learning-enhanced optical monitoring for identifying pitting-susceptible zones in 316L stainless steel. *Corros. Sci.* **256,** 113184 (2025). https://doi.org/10.1016/j.corsci.2025.113184
26. Soccol, D. & Saeys, W. Quantification of pitting corrosion resistance in stainless steel with hyperspectral imaging in the Vis/NIR range. *Corros. Sci.* **233,** 112073 (2024). https://doi.org/10.1016/j.corsci.2024.112073
27. Zhao, L., Locke, J., Xu, F., Yao, T. & Guo, X. Accurate segmentation of localized corrosion in structural alloys via deep learning. *npj Mater. Degrad.* **9,** 80 (2025). https://doi.org/10.1038/s41529-025-00633-3
28. He, K., Gkioxari, G., Dollar, P. & Girshick, R. Mask R-CNN. In *2017 IEEE International Conference on Computer Vision (ICCV)* 2980-2988 (2017). https://doi.org/10.1109/iccv.2017.322
29. Bolya, D., Zhou, C., Xiao, F. & Lee, Y. J. YOLACT: Real-Time Instance Segmentation. In *2019 IEEE/CVF International Conference on Computer Vision (ICCV)* 9156-9165 (2019). https://doi.org/10.1109/iccv.2019.00925
30. Carion, N. *et al.* End-to-End Object Detection with Transformers. In *Lecture Notes in Computer Science* 213-229 (2020). https://doi.org/10.1007/978-3-030-58452-8_13
31. Kirillov, A. *et al.* Segment Anything. In *2023 IEEE/CVF International Conference on Computer Vision (ICCV)* 3992-4003 (2023). https://doi.org/10.1109/iccv51070.2023.00371
32. Xiao, B. *et al.* Florence-2: Advancing a Unified Representation for a Variety of Vision Tasks. In *2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)* 4818-4829 (2024). https://doi.org/10.1109/cvpr52733.2024.00461
33. Minderer, M., Gritsenko, A. & Houlsby, N. Scaling Open-Vocabulary Object Detection. *Adv. Neural Inf. Process. Syst.* **36,** 72983-73007 (2023). https://doi.org/10.52202/075280-3191
34. Liu, S. *et al.* Grounding DINO: Marrying DINO with Grounded Pre-training for Open-Set Object Detection. In *Lecture Notes in Computer Science* 38-55 (2025). https://doi.org/10.1007/978-3-031-72970-6_3
35. Nash, W., Zheng, L. & Birbilis, N. Deep learning corrosion detection with confidence. *npj Mater. Degrad.* **6,** 26 (2022). https://doi.org/10.1038/s41529-022-00232-6
36. Maier-Hein, L. *et al.* Metrics reloaded: recommendations for image analysis validation. *Nat. Methods* **21,** 195-212 (2024). https://doi.org/10.1038/s41592-023-02151-z
37. Reinke, A. *et al.* Understanding metric-related pitfalls in image analysis validation. *Nat. Methods* **21,** 182-194 (2024). https://doi.org/10.1038/s41592-023-02150-0
38. Xue, W. *et al.* Initial localized corrosion induced by multiscale precipitates in the new generation high-strength Al-Zn-Mg-Cu alloy. *Corros. Sci.* **224,** 111516 (2023). https://doi.org/10.1016/j.corsci.2023.111516
39. Rafailov, R. *et al.* Direct Preference Optimization: Your Language Model is Secretly a Reward Model. *Adv. Neural Inf. Process. Syst.* **36,** 53728-53741 (2023). https://doi.org/10.52202/075280-2338
40. Meng, Y., Xia, M. & Chen, D. SimPO: Simple Preference Optimization with a Reference-Free Reward. *Adv. Neural Inf. Process. Syst.* **37,** 124198-124235 (2024). https://doi.org/10.52202/079017-3946

41. Barbey-Binggeli, M. & Tileli, V. Pit Morphology, Dissolution Kinetics, and Gas Generation Monitored in Real Time during Localized Anodic Aluminum Corrosion. *J. Am. Chem. Soc.* **147,** 34920-34932 (2025). https://doi.org/10.1021/jacs.5c11352
42. Kapoor, S. & Narayanan, A. Leakage and the reproducibility crisis in machine-learning-based science. *Patterns* **4,** 100804 (2023). https://doi.org/10.1016/j.patter.2023.100804
43. Soille, P. *Morphological Image Analysis* (Springer, 2004). https://doi.org/10.1007/978-3-662-05088-0
44. Loshchilov, I. & Hutter, F. Decoupled weight decay regularization. In *International Conference on Learning Representations* (2019). https://arxiv.org/abs/1711.05101
45. Otsu, N. A Threshold Selection Method from Gray-Level Histograms. *IEEE Trans. Syst. Man Cybern.* **9,** 62-66 (1979). https://doi.org/10.1109/tsmc.1979.4310076

Acknowledgements

This work was supported by the National Natural Science Foundation of China (grant no. 52272446), the Natural Science Foundation of Shaanxi Province (grant no. 2025JC-YBQN-654), and the Fundamental Frontier Projects (grant no. 2025-QYCX-YB-07-039).

Author contributions

H.L. designed the algorithms, constructed the dataset and conducted the experiments. D.L., X.Y. and B.R. provided guidance on manuscript writing and experimental work. G.T. provided guidance on the research direction and experimental work.

Competing interests

The authors declare no competing interests.