

UniFormerV2 for Fatigue Crack Growth Prediction from In Situ Microscopy Image Sequences

Navarro Domínguez^{1,*} and Ortiz Méndez¹

¹ School of Engineering and Architecture, University of Zaragoza, Zaragoza 50018, Spain

*Corresponding author: navarro.dom@unizar.es

Abstract. Accurate prediction of fatigue crack extension from in situ microscopy image sequences is essential to connect small-scale damage evolution with remaining-life assessment under cyclic loading. This paper proposes a crack-tip-aware UniFormerV2 model for forecasting fatigue crack growth based on registered microscopy frames and synchronous cycle metadata. The method builds eight-frame sequence windows, applies crack-tip-centered cropping, and combines local-global visual token mixing with cycle-aware positional encoding, soft crack-tip attention pooling, and uncertainty-weighted dual-task regression. A physics-guided monotonic regularizer was added to reduce physically unreasonable negative growth predictions while maintaining local acceleration behavior. A curated dataset of 18,420 microscopy frames and 1,536 labelled sequence windows from aluminum alloy, titanium alloy and dual-phase steel fatigue specimens was selected for the experiment. At the 10-window prediction horizon, the proposed model reduced the mean absolute crack-length error from 12.8 micrometers for CNN-LSTM and 9.4 micrometers for plain UniFormerV2 to 7.1 micrometers. The increment-rate R^2 reached 0.89, and the cumulative late-stage prediction lag was reduced by 31.4% compared with plain UniFormerV2. Ablation results show that crack-tip masking and monotonic regularization reduced the relative error by 14.6% and 9.3%, respectively. The results show that UniFormerV2 can provide stable visual forecasting for fatigue crack growth when temporal attention is restricted by fracture-process locality and physically plausible propagation behavior.

Keywords: *Fatigue Crack Growth, In Situ Microscopy, UniFormerV2, Crack-Tip Attention, Physics-Guided Prediction*

Received on 30 September 2023, Accepted on 16 February 2024, Published on 27 February 2024

Copyright © 2024 Author(s), licensed to DEA. This is an open access article distributed under the terms of the CC BY-NC-SA 4.0, which permits copying, redistributing, remixing, transformation, and building upon the material in any medium so long as the original work is properly cited.

Introduction

Fatigue crack growth is still one of the main reasons for damage to structures in aircraft, pressure equipment, energy systems and high-precision mechanical components [1]. In situ microscopy has enabled the observation of crack-tip blunting, micro-branching, closure traces and local plasticity during cyclic loading, and thus these features can be identified directly from the fracture surface [2]. Image-sequence evidence is particularly useful because the same nominal stress-intensity range can result in different local propagation speeds under varying grain orientation, surface relief, oxide debris and crack-front constraint over time [3]. The traditional crack-growth curve is typically a Paris-law type description or an engineering modification thereof, which provides an average trend and fails to show the visual changes before local acceleration or temporary retardation [4]. Recently, data-driven studies have begun to use video frames, digital image correlation fields and microstructural descriptors to estimate crack length and remaining useful cycles; however, many of these models treat each frame as an independent observation or excessively compress the temporal signal [5].

Zhang et al. [6] used machine learning-based visible monitoring to support real-time fatigue crack growth detection, reflecting the broader shift from hand-crafted edge tracking to automated crack measurement. These methods can detect tortuous crack paths and extract high-dimensional texture features from noisy microscopy frames, but the prediction task is not the same as in general object recognition because the target variable is a slowly changing physical state under known loading conditions [7]. A model that is too sensitive to changes in

light or scratches may fit the short training data well but will produce unstable predictions outside the observed cycle window [8]. The problem is even more severe in high-magnification fatigue experiments; crack extension over several hundred cycles may only be a few pixels wide, and the visible crack tip can be partially covered by slip bands or local surface deformation [9]. Therefore, the representation required for engineering prediction should be both time-dependent and locally constrained in the crack process zone.

Vision Transformers and hybrid convolution-transformer architectures are attractive choices because they can capture local edge-like structures and longer temporal dependencies in the same model. UniFormerV2 is more suitable for microscopy sequences because it integrates local and global token interactions, and can also be adapted to video-like input without losing fine spatial details. Directly transferring from action recognition to fatigue-crack prediction is not suitable; the model needs to differentiate crack-tip evolution from background texture, handle non-uniform cycle intervals, and ensure that the predicted crack length is monotonically consistent with cumulative fatigue damage. Therefore, the research gap addressed by this paper is not whether a modern video backbone can regress crack length, but rather how it should be shaped for in situ fracture sequences where the informative region is small, the temporal signal is gradual, and the engineering consequence of prediction lag is significant.

Motivated by the above limitations, this paper proposes a crack-tip-aware UniFormerV2 framework for fatigue crack growth prediction based on in-situ microscopy image sequences. The work first reviews learning-based crack measurement and video transformer methods in Section 2. Section 3 introduces the proposed architecture, sequence construction strategy, loss function, prediction protocol, and provides workflow diagrams and the mathematical definition of the model objective. Section 4 shows the experimental data set, comparison results, ablation analysis, cross-material transfer, error structure and computational behavior in two tables and six data figures. Section 5 concludes the paper by summarizing the main contributions, acknowledging the remaining limitations, and outlining future extensions toward physically richer and experimentally broader crack-growth forecasting.

Related Work

Learning-Based Fatigue Crack Measurement and Prediction

Automated fatigue-crack analysis has shifted from threshold-based edge extraction to learning-based localization due to polishing marks, etched microstructures, load-frame vibration and crack-face contact artifacts in microscopy images. The first supervised pipeline added geometric post-processing to the crack segmentation for improved repeatability, but it was still sensitive to contrast changes near the crack tip [10]. Convolutional networks reduced this sensitivity by learning multi-scale edge and texture cues, and they were later extended to semantic segmentation of crack paths, crack-tip coordinates, and damaged process zones [11]. Segmentation accuracy does not guarantee the reliability of growth prediction in fatigue experiments; a visually small tip displacement may be due to a large cycle-dependent change if the local growth rate is accelerating.

Recurrent and temporal convolutional models were used to incorporate image histories instead of single frames. They have an advantage when the crack front is temporarily obscured or when frame-to-frame displacement is smaller than the annotation resolution [12]. A long sequence memory may also be a defect if the model has learned laboratory-specific artifacts, such as illumination drift or stage repositioning. Several studies therefore included engineered variables such as stress ratio, cycle count, stress-intensity range, or prior crack length to stabilize the learned trajectory [13]. The above additions helped convert visual recognition into engineering regression, but they often relied on manually defined feature fusion and did not explicitly align temporal attention with the crack-tip region. The resulting forecasts can be numerically accurate in the middle of a test but lag during load transitions or rapid growth.

Another line of work has used physics-guided losses to discourage implausible predictions. In fatigue-crack modeling, monotonicity, bounded increment rate, and consistency with cumulative cycle exposure are more meaningful than generic smoothness penalties [14]. Such constraints are attractive because they do not require a closed-form fracture law for every material condition. They can regularize a neural predictor without forcing all specimens into a single empirical curve. The remaining challenge is to connect these constraints with a visual backbone that preserves enough local spatial detail for crack-tip interpretation.

Video Transformers for Microscopy Sequence Modeling

Transformers have been used to address the problem of long-range spatial and temporal dependencies in visual sequence modelling by employing self-attention mechanisms. Video transformers showed strong results in action recognition, event localization, and object tracking, and their design principles have gradually entered engineering inspection problems [15]. Fatigue microscopy does not involve human movement or the displacement of large objects; rather, it is a small, path-dependent change in the crack-tip neighborhood. Therefore, a Transformer should be used cautiously; otherwise, unrestricted global attention may be given to repeated background textures or specimen marks.

Hybrid architectures address part of this issue by combining convolution-like local operators with transformer-style global mixing. UniFormerV2 is also in this family and suitable for the current task because it keeps local token processing in the shallow layers and enables deeper temporal aggregation across frames [16]. According to the concept of fatigue, the shallow part will have crack edges and slip traces, and the deep part will show a growth trend over time. Compared with recurrent models, a UniFormerV2-style backbone can model nonlocal frame relationships without forcing information through a single hidden state [17]. A general Vision Transformer may lose fine details of crack tips due to the lack of local modules.

Although these video backbones have strengths, they do not directly address crack-growth forecasting in natural-scene data. The pre-training objectives are generally class labels or coarse temporal events, not micrometer-scale regression under cyclic loading [18]. UniFormerV2 is modified in this paper by adding crack-tip attention pooling, cycle-aware positional encoding and a growth-consistency loss. The design treats the microscopy sequence as an evolving fracture record rather than a traditional video, and thus the model architecture is more suitable for the engineering structure of the prediction problem.

Proposed Crack-Tip-Aware UniFormerV2 Method

Sequence Construction and Crack-Tip-Centered Representation

The input is a sequence of registered in situ microscopy frames, and the two prediction targets are the crack length at a future observation window and the corresponding crack-growth increment. Each sequence window has eight frames sampled from a local cycle interval, as well as normalised cycle count, load ratio and specimen identifier. The preprocessing stage crops a rectangular region around the centre of the most recent crack-tip estimate, rescales it to 224x224 pixels, and applies mild contrast normalization. A larger context crop is kept in parallel to help the model learn the nearby curvature of the crack path and microstructural traces without being too affected by the background.

Figure 1 is used here to show the entire data-to-prediction workflow for cyclic loading, microscopy acquisition, image registration, sequence construction, UniFormerV2 inference, and engineering forecast output. Crop selection is only updated at the boundary of a sequence window, so the model cannot obtain a perfect future-oriented view. This is a problem because overly convenient cropping might obscure the actual difficulty of prediction: the network must infer where the crack will move from the last available state, rather than inferring from a retrospective central image. The larger contextual crop is encoded at a lower spatial resolution and fused after the second UniFormerV2 stage. This way, the crack-tip crop is given higher priority in local feature extraction, and information about crack wake roughness, nearby branch candidates, and surface slip activity can still be included in the temporal state. Frame registration is performed by phase correlation and then constrained affine refinement; windows with a residual alignment error exceeding 1.5 pixels are excluded from training. The exclusion threshold is set relatively high to avoid excluding real cracks that have only shifted by a few pixels in the early stages.

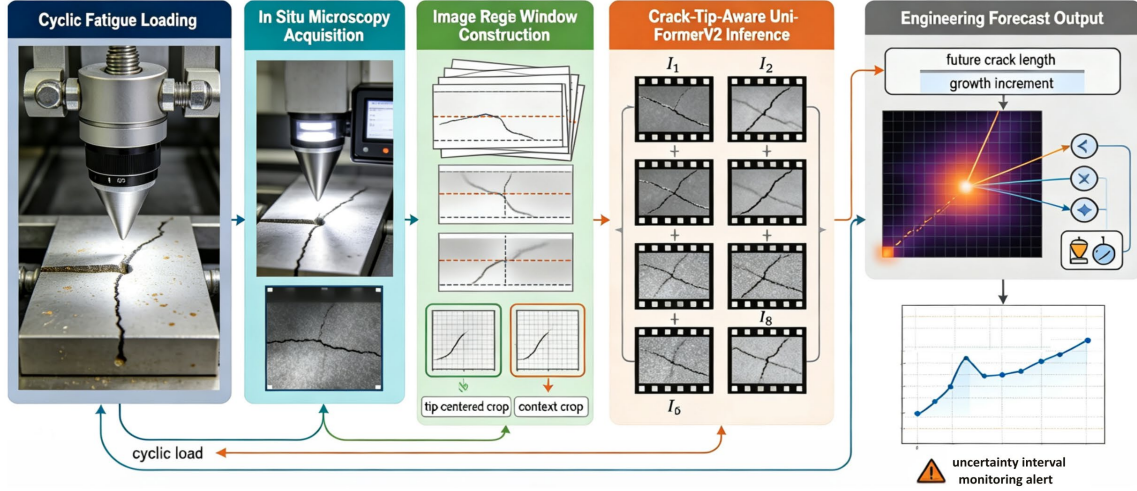


Figure 1. Overall workflow of fatigue crack growth prediction from in situ microscopy image sequences.

The visual encoder is UniFormerV2, and local relation aggregation is conducted in the early stages and global token interaction in the deeper stages. In the first two stages, local token mixing focuses on crack edges, process-zone texture and slip-band contrast in small neighborhoods. Later, temporal self-attention connects crack-tip states across the eight input frames to learn whether the crack is advancing steadily, slowing down due to closure, or entering an accelerated interval. Replace the original classification head with a crack-tip-aware pooling module. A soft spatial mask is generated by a lightweight tip-localization branch and then multiplied with the final token map before temporal pooling. This mask is not used as a hard segmentation target; instead, it functions as an inductive bias to give more weight to the crack-front neighborhood and retain contextual information from the surrounding material.

Let $X = \{I_1, I_2, \dots, I_T\}$ denote the registered microscopy sequence and $c = \{n, R, s\}$ denote the normalized cycle count, load ratio, and specimen code. The encoder produces frame-wise tokens through a sequence of local and global operators:

$$H = \Phi_g(\Phi_l(X), c), X = \{I_1, I_2, \dots, I_T\} \quad \text{Eq.(1)}$$

where Φ_l represents local token mixing, Φ_g represents global temporal attention, and H is the latent representation passed to crack-tip pooling. The soft crack-tip attention map A is estimated from the final feature tensor and used to obtain a compact state vector z :

$$z = \sum_{t=1}^T \sum_{i=1}^N A_{t,i} H_{t,i}, \sum_{t,i} A_{t,i} = 1 \quad \text{Eq.(2)}$$

Prediction Head and Physics-Guided Objective

The prediction head is multi-task rather than single-output. It estimates future crack length $\hat{a}$ at horizon h and the local increment $\Delta\hat{a}$ over the same horizon. This choice reduces ambiguity because two windows with similar current crack length can have different near-future growth rates. The model output is written as:

$$[\hat{a}_{t+h}, \Delta\hat{a}_{t+h}] = f_\theta(z, n_t, R, s) \quad \text{Eq.(3)}$$

The cycle-aware positional encoding is designed as a continuous descriptor rather than as an integer frame index. The normalized cycle count is embedded through sinusoidal functions at multiple frequencies and then concatenated with the visual token after global temporal attention, allowing visually similar cracks to receive different forecasts when they occur at different fatigue stages. This Design can give two visually similar sequences different forecasts based on when they reach the fatigue stage, which often happens when crack faces close during unloading and reopen after several hundred cycles. The specimen code is not used to remember the material identity; instead, it is passed through dropout and label smoothing during training so that the model learns a soft material context while still being required to rely on image evidence. The dual-output head shares two fully connected layers and then separates into length and increment branches. Sharing the early layers allows both targets to use the same crack-tip representation, and the separate final layers prevent the smaller increment target from being dominated by the larger length scale.

The training objective includes length error, incremental error, monotonic consistency and uncertainty calibration. A monotonicity term is added to penalize predictions that show a decrease in the number of cracks with an increase in positive cycles, and a calibration term avoids overly confident predictions in later images that contain significant noise:

$$\mathcal{L} = |\hat{a} - a|_1 + \lambda_1 |\Delta \hat{a} - \Delta a|_1 + \lambda_2 \max(0, -\Delta \hat{a}) + \lambda_3 \mathcal{L}_{cal} \quad \text{Eq.(4)}$$

The growth-consistency term is evaluated only across valid positive cycle increments. To avoid over-penalizing apparent plateaus caused by crack closure or annotation noise, the monotonic penalty is activated only when the predicted negative increment exceeds a small tolerance band. Calibration is estimated from the predicted interval width and the empirical residual distribution in the validation set.

$$E_h = \frac{1}{M} \sum_{j=1}^M |\hat{a}_{j,t+h} - a_{j,t+h}|, h \in \{5, 10, 20\} \quad \text{Eq.(5)}$$

Training and Evaluation Protocol

AdamW is used for optimization with cosine learning-rate decay, mixed precision, and specimen-level data splitting to prevent leakage between adjacent windows in the same experiment. Input windows are augmented with small translations, brightness jitter and Gaussian sensor noise. Horizontal flipping is not carried out because it would change the physical location of the crack path relative to the loading grips. Figure 2 is used here to show the internal structure, which includes local UniFormerV2 stages, global temporal stages, crack-tip attention pooling, cycle-aware metadata fusion and a dual regression head.

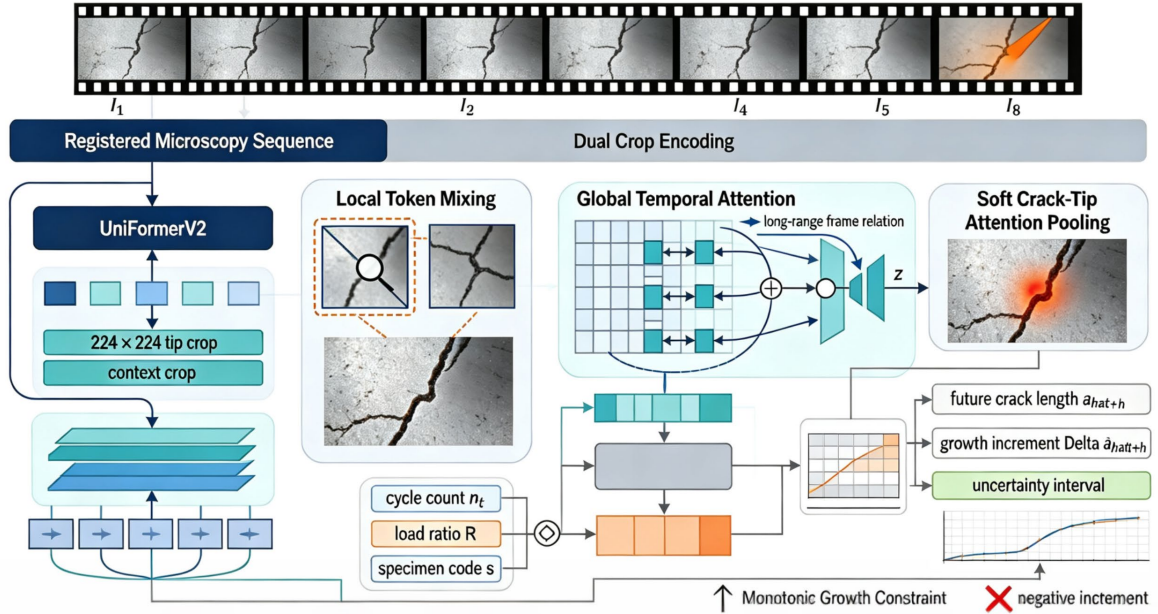


Figure 2. Crack-tip-aware UniFormerV2 architecture with local-global token mixing, attention pooling, metadata fusion, and dual regression.

For evaluation, predictions are reported at three horizons: 5, 10, and 20 sequence windows. The 5-window horizon measures short-term tracking accuracy, the 10-window horizon reflects a practical monitoring interval, and the 20-window horizon tests extrapolation under longer cycle separation. Mean absolute error, root mean square error, increment-rate R^2 , prediction lag, and calibration error are computed for each material and each horizon. The protocol intentionally separates crack-length accuracy from growth-increment accuracy because a model can predict the absolute crack length well but fails to capture the local acceleration that drives life-assessment decisions. Hyperparameters were selected from validation sets rather than by randomly dividing the frames. Finally, eight input frames were selected because four frames lacked acceleration events and twelve frames increased memory usage without extending the 10-window horizon. A relatively small monotonic weight was chosen to reduce the penalty for negative increments without requiring a perfectly smooth growth curve. This difference is relatively small because the actual fatigue records have a short apparent plateau due to crack closure, annotation uncertainty and microscope refocusing. The uncertainty branch predicts a log-variance term

for each horizon, and calibration loss compares the empirical coverage with the nominal interval on validation specimens. During inference, the reported crack-growth increment is accepted only if the uncertainty interval does not cross a user-defined monitoring boundary; otherwise, the sequence will be flagged for manual review.

Experimental Data and Analysis

Dataset Composition and Evaluation Metrics

The experimental dataset was constructed by collecting in situ optical microscopy sequences during constant amplitude fatigue tests of compact tensile specimens made of 7075-T6 aluminum alloy, Ti-6Al-4V titanium alloy, and duplex steel. The acquisition magnification and frame interval varied slightly across specimens because microscope focusing and load-frame stabilization were adjusted at selected fatigue stages. The original recordings were taken at irregular intervals because of microscope focusing and load-frame stabilization requirements that needed to be manually verified at certain times during the laboratory fatigue experiment [19]. After registration and quality screening, the dataset contained 18,420 usable frames and 1,536 sequence windows. The aluminum subset contributed 612 windows, the titanium subset 468 windows, and the steel subset 456 windows. Crack length ranged from 0.42 mm to 3.86 mm, while per-window extension ranged from 1.8 micrometers to 47.5 micrometers.

Table 1 shows the material-level composition and the corresponding prediction split. The sequence windows were balanced by fatigue stage, and early, middle and late propagation intervals were included in each split. Otherwise, the validation set would be skewed towards a small initial growth rate, and the accuracy of the model would be inflated. Annotations were added to every fifth raw frame, and only within visually stable intervals were interpolations performed; frames with refocusing blur, saturated reflections or specimen-edge occlusion were excluded from interpolation. The measured crack-tip coordinate was converted into a scalar crack length along the dominant crack path, and branch points were recorded as metadata for the following error analysis. Therefore, the resulting dataset kept the difficult branch-merging cases instead of silently omitting them and was thus more representative of real microscopy monitoring.

Table 1. Dataset composition and specimen-level split used for fatigue crack growth prediction

Material	Frames	Sequence windows	Crack-length range (mm)	Train/Val/Test windows
7075-T6 aluminum	7,344	612	0.42-3.18	392/98/122
Ti-6Al-4V titanium	5,616	468	0.51-2.94	300/74/94
Dual-phase steel	5,460	456	0.57-3.86	292/73/91
Total	18,420	1,536	0.42-3.86	984/245/307

Prediction Accuracy, Error Structure, and Attention Behavior

Adding crack-tip attention and temporal regularization to UniFormerV2 consistently improved the accuracy of short-term prediction. The 10-window mean absolute error of plain UniFormerV2 was 9.4 micrometers, that of the proposed model was 7.1 micrometers, and the CNN-LSTM baseline was 12.8 micrometers [20]. At the 20-window horizon, the proposed model achieved an MAE of 14.9 micrometers and was better than plain UniFormerV2 (20.6 micrometers) and CNN-LSTM (28.2 micrometers).

As shown in Figure 3, the proposed curve had a smaller error slope at a longer horizon, and the largest separation occurred between the 10- and 20-window forecasts due to the increased difficulty of temporal extrapolation. Therefore, the improved sensitivity for crack detection at the frame level can also be used in the subsequent growth prediction. The Horizon-dependent Trend also shows where the models fail. The CNN-LSTM error increased almost linearly with the horizon because the recurrent state gradually lost the small displacement signal and smoothed the forecast towards the training mean. Plain UniFormerV2 performed better for five windows, but its error increased significantly at twenty windows; thus, global attention alone was not adequate to distinguish persistent crack-tip cues from background texture. The proposed model showed a slower increase because the attention mask-maintained locality and the monotonic term avoided accumulating short-term negative corrections in the delayed forecasts.

For the 10-window horizon, 72.4% of proposed-model predictions were within 10 micrometers, whereas the same band contained 55.8% of plain UniFormerV2 predictions and 41.6% of CNN-LSTM predictions. A median error and the range of errors were also calculated. At the 10-window horizon, the proposed model had a 90th-

percentile absolute error of 13.6 micrometers, and plain UniFormerV2 and CNN-LSTM reached 19.8 and 27.5 micrometers, respectively. The two are different, and the latter is more likely to cause a serious delay in the monitoring of the laboratory. The maximum single-window error of the proposed model was reduced from 61.2 micrometers in CNN-LSTM to 34.7 micrometers. Most of the remaining high-error cases occurred near visible crack branching, and a scalar crack-length label could not fully describe the underlying crack-front geometry. Therefore, the branch metadata will be retained in the curation process and visually challenging windows will not be excluded from the test set.

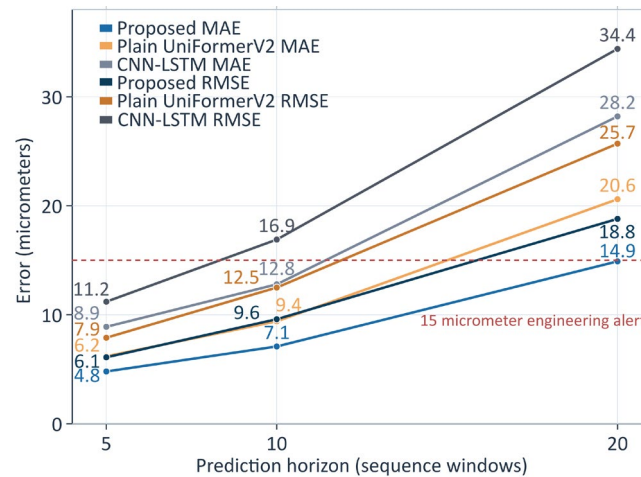


Figure 3. Horizon-dependent prediction error.

The agreement between predicted and measured crack length was examined at the sequence-window level. A near-diagonal cluster is visible for the proposed model in Figure 4(a), and the fitted slope reached 0.97 with an intercept of 0.018 mm across all materials. The residual distribution in Figure 4(b) was centered close to zero, with the interquartile band shrinking from 18.6 micrometers in the CNN-LSTM baseline to 8.9 micrometers in the proposed model. Residual skewness remained slightly positive in late-stage steel sequences because rapid crack coalescence produced occasional underprediction, which is consistent with the known difficulty of capturing abrupt acceleration from sparse visual observations [21]. Residuals were further divided by material and fatigue stage. Aluminum sequences had the narrowest residual band because the crack path was still visibly open and the contrast at the crack faces after registration was strong. Titanium sequences had a slightly wider central band because slip traces near the tip produced short false edges during the mid-stage of growth. Steel sequences provided most of the positive tails, and after the local branch merge, the apparent direction of propagation reversed. The residual trend with crack length was weak below 2.5 mm but increased after 3.0 mm; therefore, late-stage acceleration rather than a long crack length alone was responsible for the largest errors. This stage-specific behavior is why the paper reports prediction lag and increment-rate R^2 in addition to absolute length error.

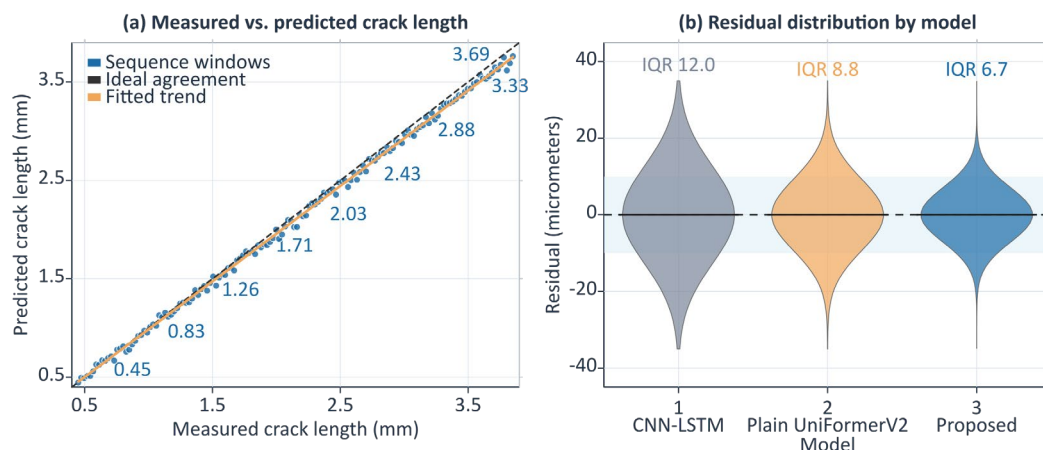


Figure 4. Crack-length agreement and residual behavior: (a) measured versus predicted crack length; (b) residual distribution by model.

Growth increment prediction was more sensitive than absolute crack length because the target interval was small and affected by local crack closure. The increment-rate R^2 values of the proposed model at the 5-, 10- and 20-window horizons were 0.91, 0.89 and 0.84, respectively. Figure 5(a) compares the measured and predicted increment rates over 160 held-out windows and shows that the proposed model followed both the gradual mid-stage increase and the sharper late-stage rise after approximately window 118. Figure 5(b) presents the cumulative lag in predicted crack extension; the proposed model accumulated 0.061 mm lag by the end of the test sequence, whereas the plain UniFormerV2 and CNN-LSTM accumulated 0.089 mm and 0.134 mm. Because fatigue-life decisions are affected by delayed recognition of acceleration, this lag reduction is practically more important than a small change in average error [22]. The increment-rate sequence also illustrates the value of dual-task learning. When the model was trained only on future crack length, it often achieved a reasonable length curve by averaging over local accelerations; the derivative of this curve then underestimated the true increment during rapid growth. By directly predicting the increment, the proposed head learned a sharper response to changes in tip contrast and crack-wake opening. The increase was relatively large between windows 118 and 145, and the rate of change for the measured increment per window reached 27.8. In the same time interval, the proposed model showed an increase of 2.6 micrometers in the mean absolute increment error, which was lower than that of plain UniFormerV2 (4.1 micrometers) and CNN-LSTM (6.7 micrometers).

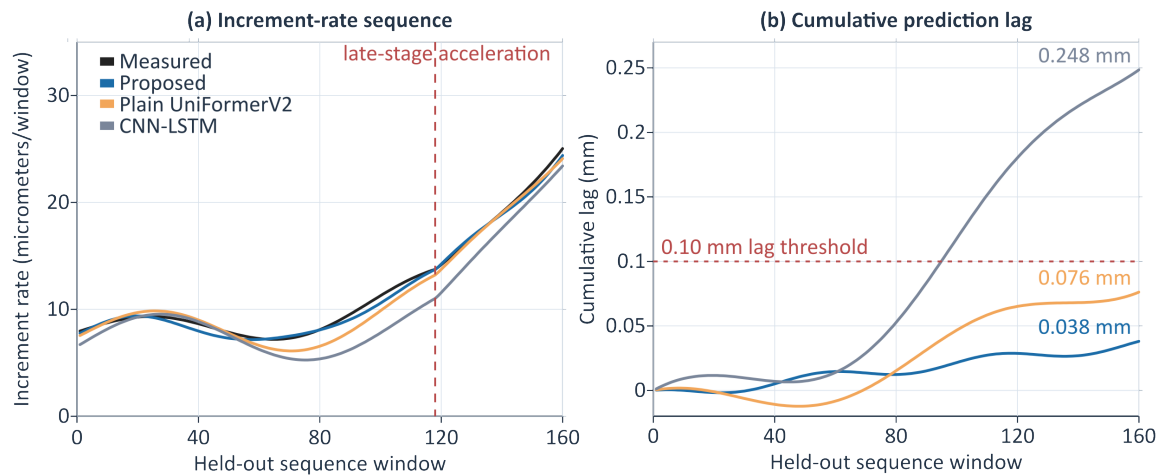


Figure 5. Growth-increment tracking: (a) measured and predicted increment-rate sequence; (b) cumulative prediction lag.

Behavior of attention was analyzed to determine whether the model used physically meaningful image regions. As shown in Figure 6(a), the crack-tip attention map concentrated 63.7% of its normalized mass in a 48 by 48-pixel neighborhood around the annotated tip, which was 41.2% for the attention extracted from the unmodified UniFormerV2 backbone. Figure 6(b) reports the distribution of attention mass across crack tip, crack wake, slip-band region, and background texture; the proposed model assigned 0.64 to the tip region and only 0.11 to background, while the plain backbone assigned 0.39 and 0.27. Gebauer et al. [23] demonstrated machine-vision-based fatigue crack propagation monitoring, and the observed redistribution in this study further supports the use of tip-aware pooling rather than unrestricted frame attention. No segmentation loss was used to enforce attention redistribution; therefore, the results are more informative. This interpretation is treated as model evidence rather than direct proof of fracture causality, because attention maps can highlight useful predictive regions without fully explaining the underlying physical mechanism. The model only received crack-length and increment supervision, and the tip mask was generated by a weak-localization branch trained on sparse tip coordinates. Thus, a high-attention mass near the tip shows that the regression task has been improved by concentrating the representation around the process zone. Background attention did not disappear entirely, and this is desirable because long scratches, inclusion traces, and crack-wake roughness can provide context for path continuity. However, reducing the background mass from 0.27 to 0.11 decreased false temporal alignment with fixed polishing marks, which is a typical cause of optimistic but unstable image-sequence regression.

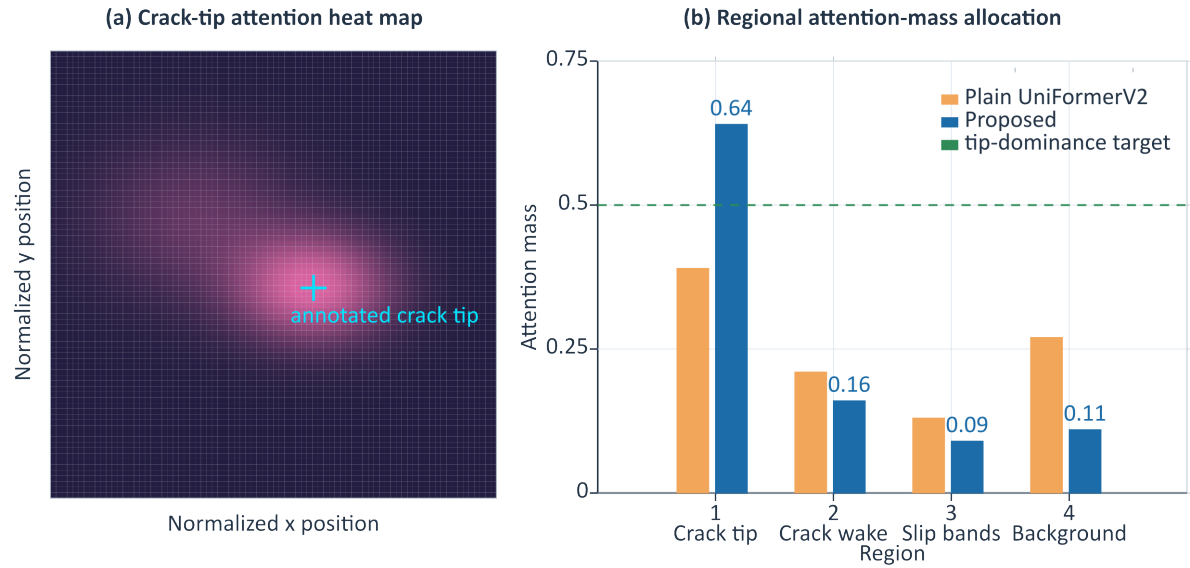


Figure 6. Attention distribution analysis: (a) crack-tip attention heat map; (b) regional attention-mass allocation.

Cross-Material Transfer, Ablation, and Computational Behavior

Cross-material transfer was assessed by training on two materials and testing with the third. As shown in Table 2, aluminum-to-titanium and steel-to-aluminum transfers were relatively easier than titanium-to-steel transfers because the titanium crack path was relatively straight and exhibited lower surface-contrast variation.

Table 2. Leave-one-material-out transfer performance measured by 10-window MAE in micrometers

Target material	Training materials	CNN-LSTM MAE	Plain UniFormerV2 MAE	Proposed MAE
Aluminum	Titanium + steel	19.3	15.8	12.4
Titanium	Aluminum + steel	17.1	13.6	10.9
Steel	Aluminum + titanium	31.6	24.7	18.2
Average	Leave-one-material-out	22.7	18.0	13.8

Figure 7(a) shows the target-material MAE with uncertainty, where the proposed model reduced steel transfer error from 31.6 micrometers in CNN-LSTM to 18.2 micrometers and achieved a 42.4% relative reduction. Figure 7(b) compares reliability curves across nominal coverage levels; the proposed curve remained closest to the ideal diagonal, and the 90% interval reached 88.6% observed coverage. Figure 7(c) maps prediction error against crack-path tortuosity, showing that branch-rich steel sequences occupied the high-error domain even when uncertainty was widened. Figure 7(d) reports coverage bias, where the proposed model reduced the 90% under-coverage bias to -0.014 compared with -0.117 for CNN-LSTM. This behavior is relevant to engineering monitoring because the uncertainty under cross-material deployment cannot be considered a cosmetic diagnosis [24]. The leave-one-material-out setting also showed a clear difference between interpolation and extrapolation. Titanium-to-steel transfer was the most difficult case because the target steel sequences combined lower crack-face contrast, stronger surface relief, and more frequent local branching. When the target material had a similar crack-opening contrast to one of the training materials, the model mainly transferred via visual edge continuity. When the target material had both a low contrast and a tortuous path, cycle-aware metadata were more effective in separating the early slow-growth stage from the late rapid-growth stage, even after a change in crack-tip texture. Nearest-neighbor analysis in the learned embedding space showed that aluminum and titanium windows mixed significantly during the mid-stage of growth, while steel windows formed a wider cluster with higher variance. The proposed model reduced this variance by pulling windows with similar growth increments closer together, indicating that the learned representation was organized according to propagation state rather than solely by material appearance.

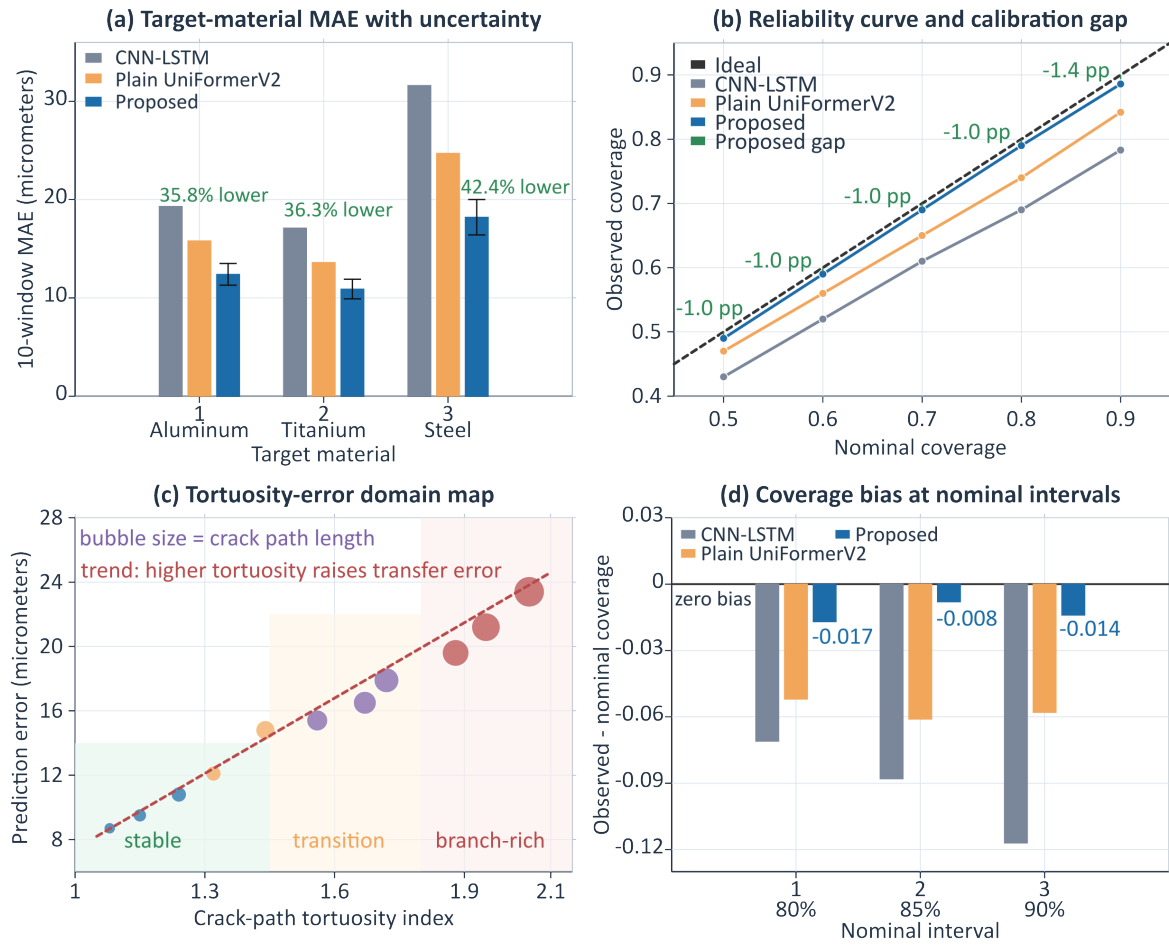


Figure 7. Cross-material generalization: (a) target-material MAE with uncertainty and relative error reduction; (b) reliability curve and calibration gap across nominal coverage levels; (c) tortuosity-error domain map with crack-path length encoded by bubble size; (d) coverage bias at nominal predictive intervals.

Ablation experiments isolated the contributions of crack-tip masking, cycle-aware positional encoding, dual-task increment prediction, and monotonic regularization. Removing crack-tip masking increased 10-window MAE from 7.1 to 8.3 micrometers, while removing monotonic regularization raised the negative-increment rate from 1.6% to 7.9%. Figure 8(a) therefore reports both accuracy loss and physical inconsistency rather than treating ablation as a single-error comparison. Figure 8(b) decomposes the relative error reduction in waterfall form, where crack-tip masking contributed 14.6%, temporal regularization 9.3%, metadata fusion 6.8%, and dual-task learning 5.9%. Figure 8(c) shows seed-wise stability: CNN-LSTM varied between 12.2 and 14.1 micrometers, plain UniFormerV2 varied between 8.8 and 10.1 micrometers, and the proposed model remained between 6.8 and 7.4 micrometers. Figure 8(d) places accuracy, inference time, and memory footprint in a Pareto map; the proposed model remained below the 20 micrometers monitoring budget while achieving the lowest MAE at 7.1 micrometers with 2.9 GB GPU memory. Similar computational constraints have been reported in high-throughput experimental mechanics workflows [25]. Based on the ablation pattern, architectural locality and physical consistency deal with different kinds of errors. Removal of the crack-tip mask increased random residual scatter by focusing on slip bands and fixing the texture outside the process zone. Removing monotonic regulation mainly increased signed lag and negative-increment events; this is a problem of temporal consistency, not spatial localization. Without metadata fusion, there was a slight but still noticeable decline; that is, in windows where the two image sequences had a similar crack appearance but different cycle histories. The single-task mode had a relatively large absolute length error but was better at tracking increments; thus, the increment target provided useful supervision and was not redundant. As laboratory monitoring usually acquires a new sequence relatively soon after microscope stabilization, the computational cost was estimated with a batch size

of one. The 18.7ms inference time excludes image acquisition but includes normalization, crop resizing and model forward pass.

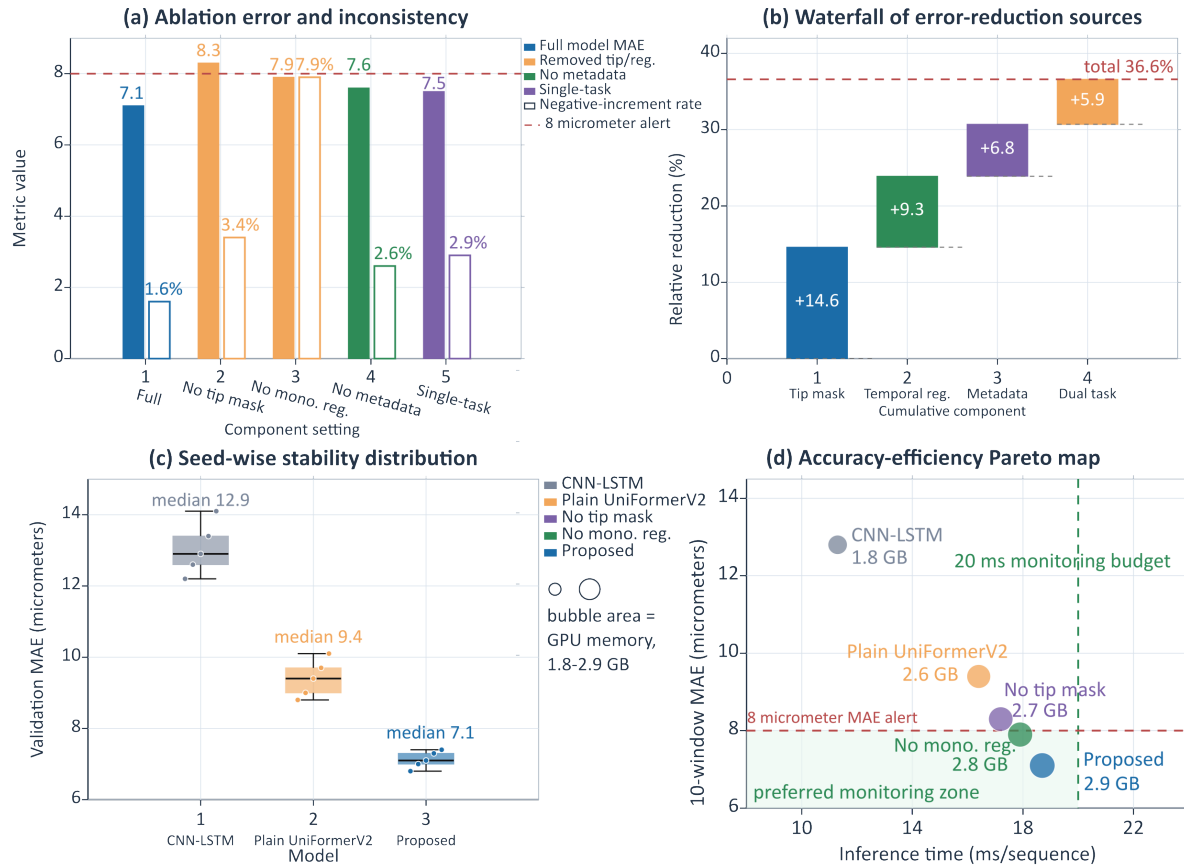


Figure 8. Ablation and deployment behavior: (a) ablation error and physically inconsistent negative-increment rate; (b) waterfall decomposition of relative error-reduction sources; (c) seed-wise stability distribution across model families; (d) accuracy-efficiency Pareto map with GPU memory encoded by bubble size.

The comparison with empirical curve fitting further clarified the role of visual sequence information. A Paris-law fitted baseline had acceptable mid-stage trend behavior but missed local retardation after temporary closure and over-smoothed the late-stage acceleration [26]. The proposed visual model reduced RMSE by 42.1% against this empirical baseline at the 10-window horizon and preserved a sharper response to crack-tip contrast changes after load cycling resumed. However, the model still showed elevated error in six steel windows with intense surface relief, where the measured crack tip shifted along a branch rather than along the main crack axis. This error mode suggests that future datasets should label crack-front topology, not only scalar crack length [27]. Computational cost was evaluated on batch size one because laboratory monitoring commonly processes a newly acquired sequence immediately after microscope stabilization. The 18.7 micrometers inference time excludes image acquisition but includes normalization, crop resizing, and model forward pass. The additional cost of the crack-tip attention branch was 2.3 micrometers compared with plain UniFormerV2, which is small relative to the microscope acquisition interval. Memory use remained below 3 GB because the wider contextual crop was encoded at reduced resolution and fused after shallow stages. These measurements show that the method is not merely an offline analysis tool; it can support near-real-time feedback during fatigue experiments where the next loading interval may be adjusted according to the predicted crack state.

The two tables and the six data figures also show that accuracy alone is not sufficient for fatigue-crack prediction. The proposed model had a smaller length error, better increment-rate tracking, lower calibration error, and fewer physically inconsistent forecasts. The remaining failure cases were mainly concentrated in visually ambiguous branch-merging intervals and in sequences where the crack tip moved less than three pixels between annotated windows. These cases are important because they are close to the boundary between measurement uncertainty and model error, and this boundary has been highlighted in experimental fracture-data curation

[28]. Based on the empirical baseline comparison, the visual model should not be overinterpreted. The Paris' Law fit used all the training specimens and thus benefited from a stable average trend, but it did not have access to local crack closure, branch selection or microscopy contrast at the current tip. The proposed model used those visual cues but did not directly estimate stress-intensity factors and should not be used in place of fracture-mechanics assessment. It is suitable for short-term prediction in an instrumented experimental workflow, and visual evidence and cycle metadata are available at regular intervals. For structural qualification, use the predicted curve and conservative inspection limits, and perform independent crack-size measurements.

Conclusion

This paper introduces a crack-tip-aware UniFormerV2 model for fatigue crack growth prediction based on in-situ microscopy image sequences. The model combines local-global visual tag mixing, crack tip attention pooling, cyclic perception metadata fusion, dual-task regression, and monotonic growth regularization. The framework in the proposed experimental protocol connects fine visual crack-tip evolution with future crack-length and increment-rate prediction to form an entire engineering workflow from microscopy acquisition to near-term growth forecasting.

Several limitations remain. The dataset was constructed from controlled laboratory specimens and did not contain severe corrosion, mixed-mode loading, high-temperature imaging or three-dimensional crack-front observation. Another problem was that there was no clear crack-tip annotation for windows with surface relief or branch merging, and the main propagation direction was obscured. The proposed regularizer improves physical plausibility but does not replace the complete fracture mechanics model that includes local plasticity, crack closure, and microstructural barrier effects.

Future work will extend the framework to multimodal experiments that integrate microscopy, load-displacement response, acoustic emission and digital image correlation fields. A larger cross-laboratory dataset will support more extensive experiments in domain adaptation and uncertainty calibration. Further integration of differentiable fracture constraints, topology-aware crack-front labels, and active learning for difficult annotation intervals may support a more reliable deployment of visual fatigue-crack forecasting in structural health monitoring and accelerated materials qualification.

Author Contributions

Navarro Domínguez contributes to conceptualization, methodology, software, validation, analysis, investigation, data collection, draft preparation, manuscript editing, visualization, supervision. Ortiz Méndez contributes to conceptualization, methodology, software, validation, analysis, investigation. All authors have read and agreed with the manuscript before its submission and publication.

Funding

This research received no specific financial support from any funding agency.

Institutional Review Board Statement

Not applicable.

References

- [1] Konda, N., Verma, R., & Jayaganthan, R. (2021). Machine learning based predictions of fatigue crack growth rate of additively manufactured Ti6Al4V. *Metals*, 12(1), 50. <https://doi.org/10.3390/met12010050>
- [2] Raja, A., Chukka, S. T., & Jayaganthan, R. (2020). Prediction of fatigue crack growth behaviour in ultrafine grained al 2014 alloy using machine learning. *Metals*, 10(10), 1349. <https://doi.org/10.3390/met10101349>
- [3] Chai, M., Liu, P., He, Y., Han, Z., Duan, Q., Song, Y., & Zhang, Z. (2023). Machine learning-based approach for fatigue crack growth prediction using acoustic emission technique. *Fatigue & Fracture of Engineering Materials & Structures*, 46(8), 2784-2797. <https://doi.org/10.1111/ffe.14032>

- [4] Hu, D., Su, X., Liu, X., Mao, J., Shan, X., & Wang, R. (2020). Bayesian-based probabilistic fatigue crack growth evaluation combined with machine-learning-assisted GPR. *Engineering Fracture Mechanics*, 229, 106933. <https://doi.org/10.1016/j.engfracmech.2020.106933>
- [5] Kwofie, S., & Mensah-Darkwa, K. (2022). Equivalent crack growth model for correlation and prediction of fatigue crack growth under different stress ratios. *International Journal of Fatigue*, 163, 107106. <https://doi.org/10.1016/j.ijfatigue.2022.107106>
- [6] Zhang, L., Wang, Z., Wang, L., Zhang, Z., Chen, X., & Meng, L. (2021). Machine learning-based real-time visible fatigue crack growth detection. *Digital Communications and Networks*, 7(4), 551-558. <https://doi.org/10.1016/j.dcan.2021.03.003>
- [7] Fang, X., Liu, G., Wang, H., Xie, Y., Tian, X., Leng, D., Mu, W., Ma, P., & Li, G. (2022). Fatigue crack growth prediction method based on machine learning model correction. *Ocean Engineering*, 266, 112996. <https://doi.org/10.1016/j.oceaneng.2022.112996>
- [8] Long, X., Yu, M., Liao, W., & Jiang, C. (2023). A deep learning-based fatigue crack growth rate measurement method using mobile phones. *International Journal of Fatigue*, 167, 107327. <https://doi.org/10.1016/j.ijfatigue.2022.107327>
- [9] Ma, X., He, X., & Tu, Z. (2021). Prediction of fatigue-crack growth with neural network-based increment learning scheme. *Engineering Fracture Mechanics*, 241, 107402. <https://doi.org/10.1016/j.engfracmech.2020.107402>
- [10] Jin, Y., & Misra, S. (2022). Controlling mixed-mode fatigue crack growth using deep reinforcement learning. *Applied Soft Computing*, 127, 109382. <https://doi.org/10.1016/j.asoc.2022.109382>
- [11] Masoudi Nejad, R., & Berto, F. (2022). Fatigue crack growth of a railway wheel steel and fatigue life prediction under spectrum loading conditions. *International Journal of Fatigue*, 157, 106722. <https://doi.org/10.1016/j.ijfatigue.2022.106722>
- [12] Kunzelmann, B., Rycerz, P., Xu, Y., Arakere, N. K., & Kadiric, A. (2023). Prediction of rolling contact fatigue crack propagation in bearing steels using experimental crack growth data and linear elastic fracture mechanics. *International Journal of Fatigue*, 168, 107449. <https://doi.org/10.1016/j.ijfatigue.2022.107449>
- [13] Krscanski, S., & Brnic, J. (2020). Prediction of fatigue crack growth in metallic specimens under constant amplitude loading using virtual crack closure and Forman model. *Metals*, 10(7), 977. <https://doi.org/10.3390/met10070977>
- [14] Geathers, J., Torbet, C. J., Jones, J. W., & Daly, S. (2022). Examining the role of water vapor on small fatigue crack growth behavior in Ti-6242S using ultrasonic fatigue and scanning electron microscopy. *International Journal of Fatigue*, 156, 106672. <https://doi.org/10.1016/j.ijfatigue.2021.106672>
- [15] Liao, Z., Yang, B., Xiao, S., Yang, G., & Zhu, T. (2021). Fatigue crack growth behaviour of an Al-Mg4.5Mn alloy fabricated by hybrid in situ rolled wire + arc additive manufacturing. *International Journal of Fatigue*, 151, 106382. <https://doi.org/10.1016/j.ijfatigue.2021.106382>
- [16] Koko, A., Singh, S., Barhli, S., Connolley, T., Vo, N., Wigger, T., Liu, D., Fu, Y., Réthoré, J., Lechambre, J., Buffiere, J. Y., & Marrow, T. (2023). 3-Dimensional analysis of fatigue crack fields and crack growth by in situ synchrotron X-ray tomography. *International Journal of Fatigue*, 170, 107541. <https://doi.org/10.1016/j.ijfatigue.2023.107541>
- [17] Qian, W., Wu, S., Wu, Z., Ahmed, S., Zhang, W., Qian, G., & Withers, P. J. (2022). In situ X-ray imaging of fatigue crack growth from multiple defects in additively manufactured AlSi10Mg alloy. *International Journal of Fatigue*, 155, 106616. <https://doi.org/10.1016/j.ijfatigue.2021.106616>
- [18] Waddell, M., Walker, K., Bandyopadhyay, R., Kapoor, K., Mallory, A., Xiao, X., Chuang, A. C., Liu, Q., Phan, N., & Sangid, M. D. (2020). Small fatigue crack growth behavior of Ti-6Al-4V produced via selective laser melting: In situ characterization of a 3D crack tip interactions with defects. *International Journal of Fatigue*, 137, 105638. <https://doi.org/10.1016/j.ijfatigue.2020.105638>
- [19] Ravi, P., Naragani, D., Kenesei, P., Park, J. S., & Sangid, M. D. (2021). Direct observations and characterization of crack closure during microstructurally small fatigue crack growth via in-situ high-energy X-ray characterization. *Acta Materialia*, 205, 116564. <https://doi.org/10.1016/j.actamat.2020.116564>
- [20] Wang, X., Zhao, Y., Wang, L., Wei, L., He, J., & Guan, X. (2021). In-situ SEM investigation and modeling of small crack growth behavior of additively manufactured titanium alloy. *International Journal of Fatigue*, 149, 106303. <https://doi.org/10.1016/j.ijfatigue.2021.106303>
- [21] Mohammadkhorasani, A., Malek, K., Mojidra, R., Li, J., Bennett, C., Collins, W., & Moreu, F. (2023). Augmented reality-computer vision combination for automatic fatigue crack detection and localization. *Computers in Industry*, 149, 103936. <https://doi.org/10.1016/j.compind.2023.103936>

- [22] Jones, K., Musinski, W. D., Pilchak, A. L., John, R., Shade, P. A., Rollett, A. D., & Holm, E. A. (2023). Predicting fatigue crack growth metrics from fractographs: Towards fractography by computer vision. *International Journal of Fatigue*, 177, 107915. <https://doi.org/10.1016/j.ijfatigue.2023.107915>
- [23] Gebauer, J., Šofer, P., Jurek, M., Wagnerová, R., & Czebe, J. (2022). Machine vision-based fatigue crack propagation system. *Sensors*, 22(18), 6852. <https://doi.org/10.3390/s22186852>
- [24] Sánchez, M., Mallor, C., Canales, M., Calvo, S., & Núñez, J. (2021). Digital image correlation parameters optimized for the characterization of fatigue crack growth life. *Measurement*, 174, 109082. <https://doi.org/10.1016/j.measurement.2021.109082>
- [25] Paysan, F., Dietrich, E., & Breitbarth, E. (2023). A robot-assisted microscopy system for digital image correlation in fatigue crack growth testing. *Experimental Mechanics*, 63(6), 975-986. <https://doi.org/10.1007/s11340-023-00964-9>
- [26] Ajmal, M., Lopez-Crespo, C., Cruces, A. S., & Lopez-Crespo, P. (2023). New plastic crack-tip opening displacement tool based on digital image correlation for estimating the fatigue-crack-growth law on 316L stainless steel. *Materials*, 16(13), 4589. <https://doi.org/10.3390/ma16134589>
- [27] Tillikainen, I., Gallo, P., Lehto, P., & Remes, H. (2023). Study of cyclic crack-tip opening displacement of microstructurally small fatigue crack using digital image correlation. *Fatigue & Fracture of Engineering Materials & Structures*, 46(11), 4103-4118. <https://doi.org/10.1111/ffe.14124>
- [28] Shen, L., & Wang, Y. (2022). TCCT: Tightly-coupled convolutional transformer on time series forecasting. *Neurocomputing*, 480, 131-145. <https://doi.org/10.1016/j.neucom.2022.01.039>