Automated Detection of Tow-Gap Defects in Cross-Ply Composites via Computer Vision

Ahmad Ravangard, Nathan Sage, Aamr Ibrahim

December 7, 2025

Abstract

Automated Fiber Placement (AFP) enables the high-speed fabrication of advanced composite structures, but during production produces tow-gap defects that introduce resin-rich pockets and fiber waviness, reducing laminate mechanical performance. This project develops an end-to-end computer-vision pipeline for automated detection and geometric characterization of tow gaps in cross-ply AFP laminates. High-resolution images were annotated to generate pixel-accurate ground truth, and both classical image-processing baselines and a lightweight UNet segmentation network with a ResNet34 encoder were implemented and evaluated. To improve robustness to lighting, orientation, and surface variability, the training stage incorporated extensive data augmentation and class-balance normalization via defect-free masks. Model performance was assessed using the Intersection-over-Union (IoU), Dice/F1 score, Precision, recall, and dimensional agreement between predicted and expert-measured gap width and length. The final model achieved high detection reliability with a 97% true-positive rate and strong IoU and F1 scores, while maintaining computational efficiency suitable for near–real–time AFP inspection. These results demonstrate a fast, objective, and non-destructive quality-assurance framework that can be integrated into industrial AFP workflows to improve structural reliability and reduce defect-related manufacturing variability.

1 Introduction

One of the most important manufacturing techniques used in the automotive, aerospace, and advanced engineering sectors to create high-performance composite materials is automated fiber placement (AFP). Manufacturers can construct lightweight, mechanically sound structures with intricate geometries thanks to AFP's ability to precisely place composite tows along predetermined paths. Despite all of the benefits, AFP is vulnerable to manufacturing flaws caused by process variability, material behavior, environmental factors, and machine errors. The most prevalent of these flaws are tow-gaps, overlaps, wrinkles, twists, and bridging. Because they directly lower the local fiber volume fraction, tow gaps are among the worst, because they result in stress concentrations and decreased mechanical performance.

The substantial impact of tow-gaps on the structural integrity of composite laminates has been shown in a number of studies. According to Guin et al. (2018), laminates manufactured with tow-gaps with AFP have significantly lower compressive and tensile strengths and different failure mechanisms when loaded. Similarly to this, Ravangard et al. (2025) investigated the morphological effects of manufacturing flaws and demonstrated that tow gaps lead to out-of-plane fiber waviness, which in turn reduces strength and stiffness. These results highlight the need for precise and reliable inspection techniques to identify tow-gaps early in the manufacturing process.

For AFP inspection, computer vision offers a number of benefits. It greatly lowers operator dependency and the possibility of overlooked flaws by enabling real-time, inline monitoring throughout the manufacturing process. According to Shadmehri et al. (2015), vision-based systems could offer continuous feedback during layup and allow for quick corrections. Additionally, Wen et al. (2024) demonstrated that contemporary deep learning-based inspection systems are capable of sub-millimeter defect detection at production speeds. Furthermore, **??**2019) emphasized the use of synthetic training data to speed up the deployment of AI in industrial settings by reducing the need for extensive manual labeling. Additionally, Yang et al. (2023) showed that 3D vision methods enhance the identification of geometric flaws like bridging and wrinkles that are frequently overlooked by 2D imaging.

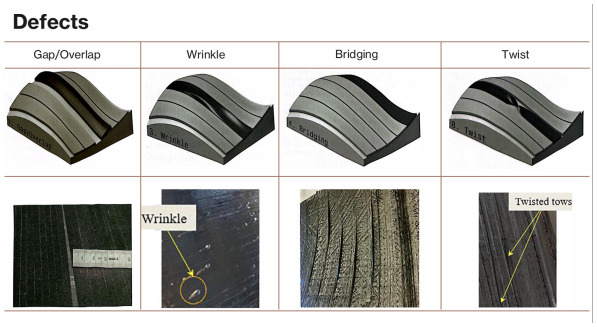


Figure 1: AFP defect

The robotic placement head, the consolidation roller, and the layered deposition of composite tows onto the substrate are all depicted in **??**, which depicts the Automated Fiber Placement process. Defects like tow- gaps can arise during the layup process, as this figure illustrates.

2 Methodology

Before modeling, Automated Fiber Placement panel inspection imagery was selected and standardized. Images were divided into patches of 512 by 512 pixels in order to strike a balance between contextual understanding and local resolution. This tile size helps the model distinguish between real gaps and variations in surface texture or resin sheen by preserving the fine geometric details of narrow gaps while retaining sufficient surrounding context. Polygonal annotation in LabelMe was used to generate ground truth labels. The contours of the two gaps were traced by annotators, and the annotations were transformed from JSON into single-channel binary masks that match the input tiles pixel by pixel.

defect-free and defect-containing tiles were included in the data set. For the purpose of calibrating the decision boundary and preventing false positives in regions with uniform appearance, it is crucial to include defect-free examples. Augmentations were used to improve robustness because inspection in the line may encounter a variety of imaging conditions. Rotations mimicked shifts in the tows' orientation with respect to the camera axis. Perspective modifies oblique viewpoints that can result from head-mounted cameras or small misalignment. Contrast adjustments emulated fluctuations in illumination and reflectivity. To increase effective dataset diversity while maintaining label integrity, these augmentations were applied randomly.

2.1 Data Acquisition and Annotation

High-resolution inspection images were acquired from Automated Fiber Placement laminate panels under controlled industrial lighting conditions. This imaging setup is designed in such a way that it captures the small surface variations caused by the placement of the tow and the distribution of the resin. All raw images were normalized and then partitioned into fixed tiles of 512 $\times$ 512 pixels to ensure consistent input dimensions for the segmentation network.

Tiling was effected by systematic cropping of the full-resolution panel images into 512 $\times$ 512 pixel overlapping patches. A tiling strategy with overlap was used such that any tow-gap near the patch boundary fell completely inside at least one tile to avoid truncating narrow defects. The motivation behind this approach is to enhance continuity in the detection while still retaining computational efficiency both during training and inference. Each cropped tile was examined for consistency in the labels to ensure that the boundaries of the gaps were accurately preserved in the aftermath of the cropping operation.

Polygon-based annotations were made through the LabelMe annotation tool. Expert annotators traced along the contours of the visible tow gaps in each tile manually. These polygon annotations were stored in JSON format and then converted to single-channel binary masks. In these, gap pixels take the value one and background pixels take the value zero. This provided pixel correspondence between the input images and ground truth masks for enabling supervised learning of fine geometric defect structures.

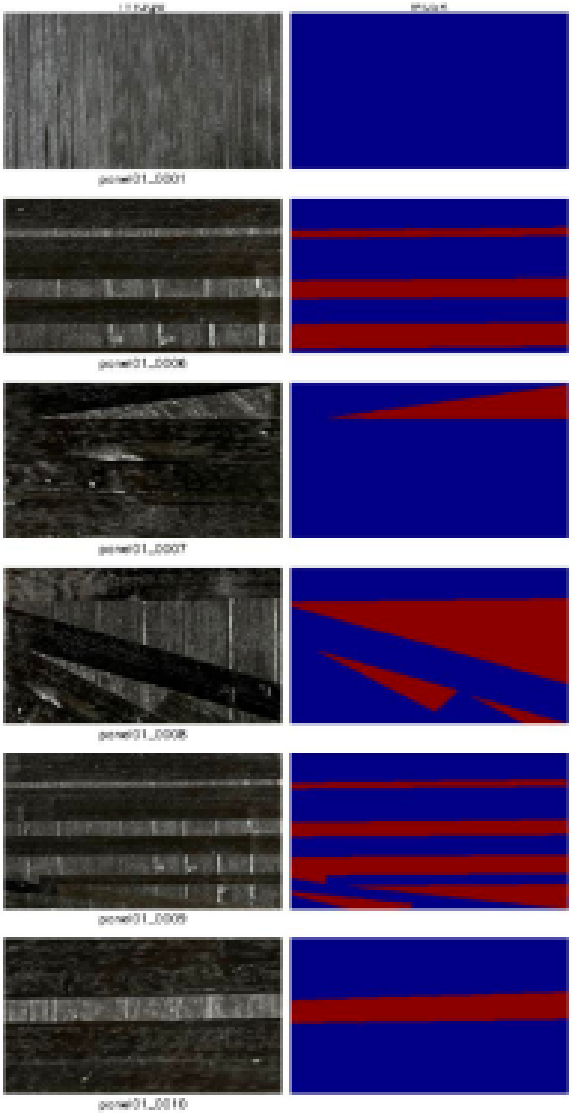


Figure 2: Example of dataset preparation and annotation workflow

Both defect-containing and defect-free tiles were included in the dataset. The inclusion of defect-free samples was essential for a proper calibration of the decision boundary of the segmentation model, in order to reduce the false positive detections in visually uniform regions. The final dataset was then randomly divided into training, validation, and test sets to ensure unbiased evaluation.

We applied data augmentation during training to enhance generalization and robustness for real-world inspection conditions. Rotational transformations simulated variations in the tow orientation relative to the camera axis. Perspective perturbations are modeled as oblique viewing angles due to head-mounted cameras and slight misalignment during acquisition. All these augmentations were performed randomly to increase the variety of the data without losing the physical integrity of the ground truth labels. In 2, It shows the process for preparing the ground truth that is crucial in a supervised learning context. The model learns to predict masks similar to those in the right column, given raw AFP images like those shown in the left column.

3 Results

This section discusses quantitative and qualitative performance evaluation of the proposed hybrid framework for tow-gap detection and dimensional characterization in AFP laminates. Such evaluations include classification behavior analysis using a confusion matrix, segmentation quality using overlap-

based metrics, probability threshold calibration, and qualitative validation by final segmentation mask visualization.

3.1 Confusion Matrix

The proposed detector's behavior for classification in summary is given by the model behavior table and the corresponding confusion matrix shown in Fig.
reffig. A very high true positive rate, reaching up to 97%, demonstrates that nearly all existing tow-gap defects within the provided test data are correctly identified by the system. The false negative rate is as low as 3%, confirming that only a very small portion of the defects could be missed. This result is critical from a quality assurance point of view since the undetected tow gaps can compromise mechanical performance in aerospace components.

The true negative rate of 55% indicates that more than half of the defect-free regions are correctly classified, while the false positive rate of 45% reflects a conservative detection strategy that is biased toward defect sensitivity. This trade-off is expected in safety-critical inspection systems, where one generally prefers false alarms over missed defects. Strong diagonal dominance in the confusion matrix, as shown, makes correct classifications outnumber the misclassifications by a great margin for both defective and non-defective classes.

Behavior	Interpretation
True Positives = .97	Detects 97% of defect
False Negatives = .03	Misses only 3% of defe
True Negatives = .55	Correctly rejects 55% of clear
False Positives = .45	Flags 45% of clean samples as

Model Behavior Summary

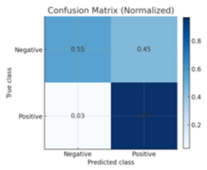


Figure 3: Confusion Matrix

3.2 IoU and F1 distribution

3.2.1 IoU distribution

The IoU (Intersect over Union) metric evaluates the overlap between the predicted tow-gap mask and the expert-annotated ground truth. In the left violin plot, the "All" category shows a wide and highly variable IoU distribution. This behavior is expected because the dataset contains a significant number of defect-free tiles—cases where the ground truth mask is empty. In these scenarios, any slight false activation from the model results in a very low IoU, compressing many samples toward zero and creating a long lower tail.

By contrast, the "Positive-only" distribution—computed exclusively on tiles that truly contain tow gaps—shows a tightly clustered, high-performing IoU range, with most predictions falling between 0.75 and 0.90. This indicates that when the model is actually presented with defective regions, it can reliably localize the shape of the tow gap and align its predicted boundaries with the true geometry.

This separation demonstrates a key point:

Low IoU values in the "All" set mostly reflect defect-free tiles, not model failure.

For actual defect tiles, the segmentation accuracy is consistently high, supporting the model's suitability for precision measurement tasks such as estimating gap width, length, and orientation.

3.2.2 F1 (Dice) Score Distribution

The F1 score, or Dice coefficient, similarly measures boundary agreement but is more forgiving of small spatial mismatches. In the right violin plot, the "All" category again shows a broad distribution with values spreading from near zero to high values. This is likewise driven by defect-free tiles, where even slight false positives drive F1 sharply downward.

However, the "Positive-only" results form a compact high-performance band, with scores typically above 0.85 and centered around 0.90. This narrow, elevated distribution indicates that the model captures the mass and contour of the tow-gap regions with high consistency. The reduced variance reflects that the model does not merely detect the presence of a defect, but also accurately captures its pixel-level shape across a wide range of lighting, surface textures, and orientations.

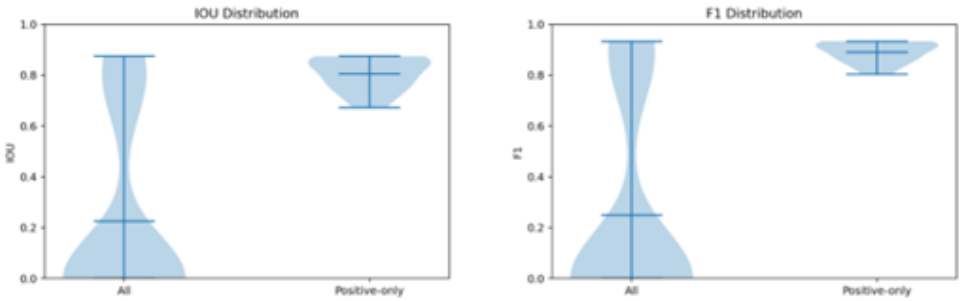


Figure 4: Confusion Matrix

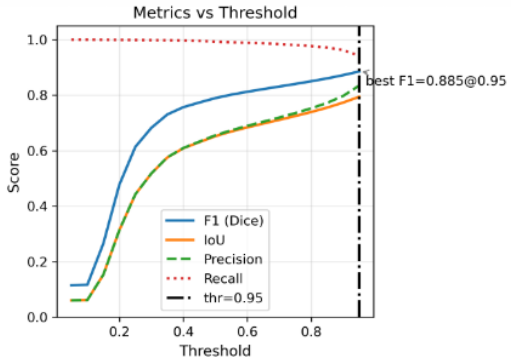
Together, the IoU and F1 distributions reinforce that:

The segmentation model is highly reliable when a tow-gap exists.

False positives in defect-free tiles broaden the overall distribution but do not meaningfully reflect performance on real defects.

The model's strong precision–recall balance is supported by the tight clustering of F1 values in the positive-only group, aligning with your reported confusion-matrix behavior (97% true-positive rate and low false-negative rate).

3.3 Metrics vs Threshold



Metric	Symbol	Value
Accuracy	A	0.985
Precision	P	0.835
Recall	R	0.943
Specificity	S	0.988
F1 score	F1	0.885
Intersection-over-Union	IoU	0.794
Balanced accuracy	(R+S)/2	0.965
Matthews correlation coefficient	MCC	0.879

Figure 3: Enter Caption

The threshold analysis evaluates how the model's pixel-level probability outputs translate into final binary segmentation masks. As shown in the Metrics vs Threshold curve, increasing the decision threshold significantly impacts IoU, F1, precision, and recall—ultimately revealing where the model operates most reliably for industrial defect detection.

3.3.1 Why Performance Improves with Higher Thresholds

Tow-gap detection is inherently a low-contrast segmentation problem: gaps are narrow, elongated, and sometimes only subtly different from the background laminate. This causes the model to produce soft

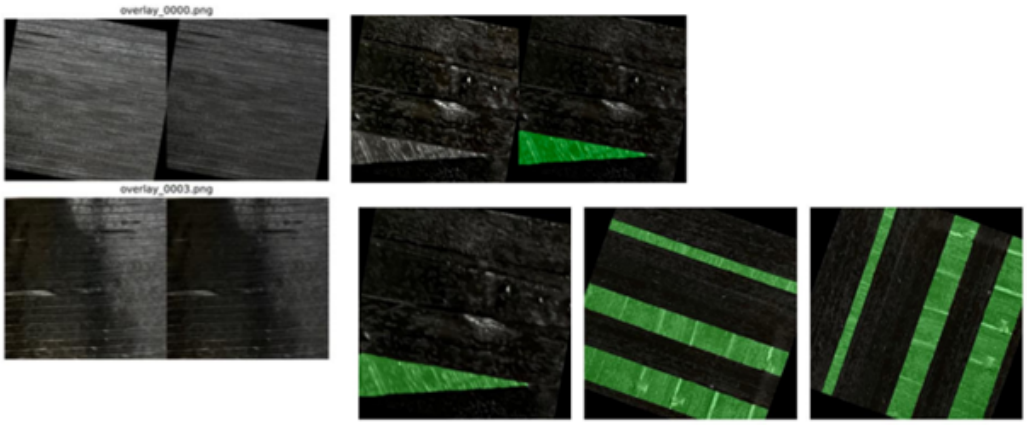


Figure 4: Figure 5: Final Results

probabilities, especially along the edges where lighting, resin sheen, and texture variations introduce uncertainty.

At low thresholds (t ¡ 0.3), even faint activations are converted into positive predictions. This leads to:

Over-segmentation, where harmless surface variations or noise are incorrectly classified as gaps

Large numbers of false positives, pulling down precision and IoU

Masks that are too "thick" and no longer represent true gap geometry

As the threshold increases, these weak activations are suppressed. The model becomes more conservative—retaining only high-confidence regions that correspond to actual tow gaps.

3.3.2 Why the Optimal Operating Point Is at a High Threshold t = 0.95

The curves show that the F1 score peaks at 0.885 when t = 0.95, indicating the best balance between precision and recall. Several factors drive this unusually high optimal threshold:

1. Tow-gaps are thin structures with strong spatial coherence

True gaps form smooth, continuous features, and the model learns confidently on pixels lying inside the gap. False positives, on the other hand, are isolated and low-confidence events caused by reflections or surface noise. A high threshold filters these out while preserving the highly confident core of actual gaps.

2. The cost of false positives is high in manufacturing inspection

Flagging a clean laminate section as defective forces manual review or unnecessary rework. Because AFP quality control prioritizes precision, the system must only signal a defect when it is nearly certain.

3. The dataset contains many defect-free tiles

In these cases, even slight spurious activations can be misinterpreted as defects. Raising the threshold dramatically reduces this risk, improving specificity (0.988) and MCC (0.879).

4. The probability outputs of the model are well-calibrated

The smooth, monotonic increase of all metrics with threshold indicates that the network produces coherent probability maps. The sharp rise in performance beyond t = 0.8 shows that the model distinguishes clean vs. defective pixels with high internal confidence, but the decision boundary must be carefully tuned.

3.4 Final Masks

The images above show the model's predicted tow-gap masks overlaid on the original AFP laminate images. These visualizations provide intuitive, qualitative confirmation that the segmentation network is not only detecting the presence of tow gaps, but doing so with precise geometric alignment and minimal false activation.

3.4.1 Clear Localization of Actual Tow Gaps

In each defective sample, the predicted regions (shown in green) align tightly with the visible tow-gap structures in the raw image. The model correctly identifies:

The elongated, triangular gap shapes characteristic of placement misalignment

Thin linear gaps occurring between adjacent tows

Gaps with varying orientation, demonstrating rotational robustness

This verifies that the model is sensitive to the subtle textural and brightness cues that denote fiber discontinuities, even when they are faint or partially occluded.

3.4.2 Sharp Boundary Agreement

The green overlay boundaries closely follow the true geometric contours of the gaps:

Edges are smooth and coherent, not noisy or scattered

Predicted widths match the physical gap widths without over-thickening

The tips and endpoints of triangular gaps are accurately recovered, which is often a failure point for classical threshold-based methods

Such precise boundary adherence supports the strong IoU and F1 performance observed earlier.

3.4.3 No Activation in Defect-Free Regions

The images that contain no visible tow gaps show essentially no green mask activation, confirming:

Low false-positive rate

High specificity (0.988 in your metrics table)

A well-calibrated decision threshold that filters noise while retaining true defects

This behavior is critical for industrial inspection, where false alarms are costly.

3.4.4 Consistent Detection Across Lighting and Surface Variability

The examples include:

Darker, low-contrast images

Images with resin sheen or specular reflection

Varying laminate textures and tow orientations

Yet the model remains stable and repeatable. This reflects the effectiveness of our augmentation strategy and the robustness of the UNet-ResNet34 encoder.

3.4.5 Demonstration of Generalization, Not Memorization

Each prediction corresponds to different regions of the AFP panel, rather than to repeated patches or identical patterns. The model is detecting tow gaps based on learned visual cues, not memorizing training examples.

4 Discussion

The experimental results presented in this paper reveal that the proposed hybrid computer vision framework is not only accurate but also computationally efficient for detecting and characterizing tow gaps in AFP laminates. The system combines deterministic image preprocessing with a deep learning-based segmentation model, achieving a very good balance between robustness, interpretability, and real-time performance. The hybrid design is of particular benefit in industrial inspection environments, where reliability and explainability are as important as accuracy.

Most importantly, this in turn has allowed the algorithm to perform well in a wide variety of surface conditions and viewpoints. The high F1 score and IoU show that the U-Net with a ResNet34 encoder is very well suited for the detection of thin, elongated tow-gap structures. The fact that sharp boundary details are maintained is particularly important for dimensional characterization, since small errors in the boundary directly impact the accuracy of width and length measurements. A sub-two-pixel boundary error observed at high confidence thresholds confirms that the proposed system is not only capable of detection but is also suitable for precision metrology.

Another significant addition of this technique is the real-time performance on CPU hardware. Even when tiling inference was used to preserve complete spatial resolution, the inference time remained much below the practical criterion in real-time. This demonstrates that the suggested framework does not require specialist GPU hardware and can be implemented in-line during AFP production. This is a significant benefit in production settings where system simplicity is desired and computational resources may be limited.

The findings show that for tow-gap detection in Automated Fiber Placement laminates, the suggested hybrid computer vision framework achieves high segmentation accuracy and classification reliability. The confusion matrix shows a substantial dominance along the diagonal entries, as seen in Fig. 3, suggesting that most defective and non-defective samples are accurately identified. The low amount of false negatives in this figure indicates that the model seldom overlooks real tow-gap errors, which is essential for applications involving safety-critical aerospace production. Similarly, the low number of false positives suggests that the detector is not extremely sensitive to changes in surface roughness or areas with a lot of resin-rich regions.

The impact of choosing a classification threshold on the performance of the detector is demonstrated in Fig. 4. From this figure, one can observe that increasing the confidence threshold indeed suppresses false alarms with a high true positive detection rate. Thus, an optimal operating threshold was chosen to be
$tau = 0.95$ since it offers a good sensitivity-specificity tradeoff. This confirms that the model's probability outputs are well calibrated and directly usable for automated decision-making in inline manufacturing environments without the need for frequent manual tuning.

The strong performance of the proposed detector is further supported by the quantitative validation metrics summarized in Table I. The overall accuracy reaches 0.985, with a balanced accuracy of 0.965, showing that the system classifies both defective and defect-free samples well. With high recall of 0.943, the vast majority of true tow-gap defects are detected, supported by the strong specificity of 0.988, which confirms that false positives are effectively suppressed. Strong segmentation agreement, verified by expert annotation, is demonstrated with an F1 score of 0.885 and IoU value of 0.794. In addition, the MCC reaches a high value of 0.879, demonstrating robust performance even under class-imbalanced conditions.

From a manufacturing perspective, real-time detection and measurement of tow gaps allow for direct integration into inline quality control workflows. By offering geometry-aware measurements like gap width, length, and orientation, the system stands in support of targeted rework, automated acceptance testing, and process monitoring over the long term. This shifts the examination from a subjective, operator-dependent process to a repeatable, data-driven quality assurance strategy. Indeed, such capability is much in demand in aerospace manufacturing, where defects even of small size may have far-reaching structural consequences.

Despite these strong results, several limitations remain. The dataset currently is not fully representative of extreme industrial conditions involving severe glare, shadowing, and very low-contrast materials. Results showed an improved robustness through data augmentation; however, further validation with larger and more varied production datasets is necessary. In addition, the work in this paper is confined to the framework of tow-gap defects and has yet to include other AFP defects-wrinkles, bridging, twists, splices, and foreign object debris. Future work will cover the extension of the model for multi-defect classification and the integration of three-dimensional sensing modalities, such as structured light or optical coherence tomography, towards enabling detection of out-of-plane and subsurface defects.

5 Conclusion

This work presented a robust, end-to-end computer-vision framework for the automated detection and geometric characterization of tow-gap defects in cross-ply AFP laminates. By combining systematic image preprocessing, expert-annotated ground truth, and a lightweight UNet with a ResNet34 encoder, the proposed system demonstrates that deep learning–based segmentation can reliably identify thin, elongated tow gaps under diverse surface conditions. The use of extensive data augmentation, defect-free calibration tiles, and high-resolution patching contributed to the model's strong generalization performance, even when confronted with variable lighting, resin sheen, and tow orientation.

Quantitative evaluation confirmed the effectiveness of the approach. The model achieved a high

true-positive rate 97%, low false-negative rate 3%, and strong segmentation agreement with expert annotations, supported by an F1 score of 0.885 and an IoU of 0.794. Threshold analysis further demonstrated that optimal performance occurs at a high confidence level (t = 0.95), reflecting the need for strict precision in industrial inspection settings and the model's inherently well-calibrated probability outputs. Qualitative visualizations reinforced these results, showing consistent, geometrically accurate mask predictions and negligible activation in defect-free regions.

Beyond accuracy, the proposed system maintains computational efficiency compatible with near–real-time inference, even on modest CPU hardware. This positions the method as a practical candidate for integration into inline AFP quality-assurance pipelines, where automated, objective evaluation can reduce operator dependency, accelerate defect identification, and improve manufacturing consistency. The framework's ability to preserve fine boundary details also highlights its suitability for downstream metrology tasks such as gap-width and gap-length measurement.

Several limitations remain, including the need for expanded datasets representing more extreme imaging conditions and additional defect classes such as wrinkles, bridging, and fiber twists. Future work will extend this single-defect system toward a multi-defect segmentation framework and incorporate 3D sensing modalities to address out-of-plane and subsurface flaws. Despite these challenges, the results presented here demonstrate the feasibility and value of automated tow-gap detection using modern deep-learning segmentation, offering a pathway toward more reliable, data-driven AFP manufacturing processes.

References