



UNIVERSITÀ
DEGLI STUDI
DI PADOVA



DIPARTIMENTO
DI INGEGNERIA
DELL'INFORMAZIONE

UNIVERSITY OF PADUA
DEPARTMENT OF INFORMATION ENGINEERING

MASTER'S DEGREE IN COMPUTER ENGINEERING

Semantic Segmentation of Carbon-Fibre Defects from Multi-Channel Photometric Imaging

MASTER CANDIDATE

Riccardo Fazzi

Student ID 2116883

SUPERVISOR

Prof. Emanuele Menegatti

ACADEMIC YEAR: 2025/2026

GRADUATION DATE: JULY 08, 2026

Abstract

Carbon-fibre reinforced materials are increasingly adopted in high-performance industrial applications, where surface defect inspection plays a critical role in ensuring product quality and structural integrity. This thesis addresses the problem of automated defect inspection in carbon-fibre textiles using multi-channel photometric imaging data acquired from the PRO-FACTOR Fscan industrial sensing system, which produces sensor-derived representations encoding physical properties of the material surface rather than conventional RGB imagery. The work makes two primary contributions. First, starting from an existing classification-oriented industrial dataset that was not fully suitable for supervised segmentation, a new dataset of carbon-fibre textile defects was acquired and curated using a controlled inspection protocol, including pixel-level annotations across multiple defect categories. Second, a systematic experimental investigation of supervised semantic segmentation pipelines is presented, addressing the specific challenges of this industrial setting, including severe data scarcity, correlated acquisitions, class imbalance, and non-standard sensor representations. A central methodological contribution of the thesis is the analysis of data leakage induced by grouped acquisition protocols, showing how standard random splitting strategies can substantially inflate performance estimates. To address this issue, leakage-safe evaluation protocols are introduced in order to obtain more reliable measures of model generalization. The work further investigates sensor-channel combinations, normalization strategies, architecture selection, backbone capacity, imbalance-aware loss formulations, and augmentation techniques, including an adaptation of CutMix for industrial defect segmentation. Experiments based on SegFormer architectures are conducted across different evaluation regimes, providing a structured analysis of the principal factors affecting segmentation performance under realistic industrial constraints.

Contents

1	Introduction	1
1.1	Problem Statement	2
1.2	From Classification to Segmentation	2
1.3	Research Objectives	4
1.4	Main Contributions	5
1.5	Thesis Organization	6
2	Background and Related Work	7
2.1	Automated visual inspection in industrial settings	7
2.2	Defect detection paradigms	8
2.2.1	Image classification	8
2.2.2	Object detection	9
2.2.3	Segmentation-based approaches	10
2.2.4	Anomaly detection and self-supervised approaches	12
2.3	Semantic segmentation models relevant to this work	13
2.3.1	SegFormer	14
2.4	Challenges in industrial defect inspection	19
2.4.1	Class imbalance and small defect regions	20
2.4.2	Limited annotated datasets	21
2.4.3	Domain shift and non-standard sensor modalities	21
2.4.4	Data leakage and correlated acquisitions	23
2.5	Related work on textile and carbon-fibre defect inspection	23
3	Dataset Analysis and Construction	26
3.1	Industrial sensing system	26
3.2	Sensor-derived representations	28
3.3	Limitations of the original dataset	30
3.4	Dataset reconstruction	31

3.4.1	Authenticity of defect samples	32
3.4.2	Annotation methodology	32
3.5	Practical channel preference during annotation	33
3.6	Dataset statistics and defect taxonomy	34
3.7	Recycled material subset	36
3.8	Qualitative inspection and annotation rationale	37
3.8.1	Missing stitch	38
3.8.2	Gap	40
3.8.3	Fuzzball	41
3.8.4	Drilled tow	42
3.8.5	Weaving error	43
3.8.6	Fuzz yarn	44
3.8.7	Distortion	45
3.8.8	Foreign object	46
3.8.9	Broken tow	47
3.9	Dataset summary	48
4	Evaluation Methodology and Data Leakage Analysis	50
4.1	Motivation for evaluation analysis	50
4.2	Dataset splitting strategies	51
4.2.1	Random sample-level splitting	51
4.2.2	Instance-grouped split	52
4.2.3	Material-based split	54
4.2.4	Conceptual generalization regimes	55
4.3	Reproduction of the classification setup	55
4.3.1	Classification setting	56
4.4	Impact of splitting strategy on evaluation	57
4.4.1	Implications for realistic deployment	58
4.5	Implications of leakage-safe evaluation	59
5	Experimental and Methodological Framework	61
5.1	Input representation design	62
5.1.1	Human versus neural network interpretability	62
5.1.2	Channel complementarity and representation selection	63
5.1.3	Frequency-based derived representations	65
5.2	Normalization and pretrained compatibility	67
5.2.1	Azimuthal channel representation	68

5.2.2	Per-sample normalization	70
5.2.3	Compatibility with RGB-pretrained backbones	71
5.3	Segmentation architecture and transfer-learning strategy	71
5.3.1	SegFormer architecture	71
5.3.2	Pretrained checkpoint adaptation	72
5.3.3	Backbone capacity variants	73
5.4	Optimization under severe data scarcity	73
5.4.1	Focal-Tversky loss formulation	73
5.4.2	Learning-rate strategies	75
5.4.3	DropPath regularization	75
5.5	Leakage-safe evaluation methodology	76
5.5.1	Monte Carlo evaluation protocol	76
5.5.2	Experimental setup	77
5.5.3	Evaluation metrics	77
5.6	Data augmentation via CutMix	79
5.7	Alternative Approaches Explored	83
5.7.1	PatchCore anomaly detection	83
5.7.2	YOLO-based instance segmentation	85
5.7.3	Cross-dataset transfer-learning exploration	86
5.8	Chapter summary	89
6	Experimental Results	90
6.1	Input Representation Analysis	91
6.1.1	Input dimensionality	91
6.1.2	Variability under leakage-safe evaluation	92
6.1.3	Effect of normalization strategies	94
6.1.4	Frequency-based representation study	95
6.1.5	Azimuthal channel mapping	95
6.2	Backbone Capacity Comparison	96
6.3	CutMix augmentation study	98
6.3.1	Quantitative results	99
6.3.2	Qualitative comparison	99
6.3.3	Interpretation of the CutMix improvements	100
6.3.4	Scope of the supervised segmentation analysis	101
6.4	Alternative approaches and exploratory experiments	101
6.4.1	Comparison with YOLO segmentation models	101
6.4.2	PatchCore anomaly detection	104

6.4.3	Cross-dataset transfer-learning experiment	105
6.5	Discussion of Failure Modes and Remaining Limitations	106
6.5.1	Data scarcity	107
6.5.2	Material variability	107
6.5.3	Small and low-contrast defects	107
6.5.4	Limits of augmentation	107
6.5.5	Overall interpretation	108
7	Conclusions and Future Work	109
	Bibliography	111

List of Figures

2.1	Conceptual illustration of image classification for industrial defect inspection. The input image is assigned a single defect label at image level, while no explicit localization of the defective region is provided.	9
2.2	Conceptual illustration of object detection for industrial defect inspection. The model predicts a bounding box around the defective region, providing approximate spatial localization without explicitly delineating the precise defect boundaries.	10
2.3	Conceptual comparison between the main segmentation paradigms. Semantic segmentation assigns a class label to every pixel in the image, without distinguishing different instances of the same class. Instance segmentation identifies and separates individual countable object instances, while background or uncountable regions are not explicitly labeled in this visualization. Panoptic segmentation combines both perspectives by assigning a semantic label to every pixel and distinguishing individual instances for countable objects. . . .	11
2.4	Illustration of semantic segmentation. Unlike image classification and object detection, the model produces a pixel-level prediction mask that explicitly captures the shape and spatial extent of the defect.	12
2.5	SegFormer encoder–decoder architecture.	14
2.6	Structure of a SegFormer block.	16
2.7	Examples of defect regions with different spatial extent. Industrial defects may range from very small localized anomalies to larger elongated structures, making robust segmentation more challenging.	21
2.8	Example of a subtle textile defect from the AITEX dataset [3]. In this case, the anomaly is difficult to identify visually because it mainly appears as a local variation in the spacing or structure of the textile pattern, rather than as a strong change in color or overall appearance. The right image highlights the annotated defect region through the ground-truth mask.	22

3.1	PROFACTOR Fscan inspection system used for data acquisition in this work.	27
3.2	Example of the raw directional observations acquired for the same material region under different illumination directions. Variations in brightness and shadow patterns emphasize different surface and fibre characteristics depending on the illumination geometry.	28
3.3	Example of the sensor-derived representations generated by the Fscan processing pipeline from the raw directional illumination observations. The azimuthal channel encodes local fibre orientation through angular values and is shown both as a normalized single-channel representation and as an HSV-based visualization used to highlight orientation differences. The remaining channels emphasize complementary physical properties of the material surface, including fibre orientation, fibre reflectivity, and reflectance behaviour under directional illumination.	29
3.4	Acquisition convention adopted for the new dataset. The same physical defect is acquired multiple times under controlled variations of position and orientation with respect to the sensor reference frame.	31
3.5	Compared with standard woven textile samples, the recycled material exhibits highly irregular fibre arrangements and reduced structural regularity, making defect delineation considerably more difficult. The first sample contains both a missing stitch (red) and a gap (green), illustrating the complex defect configurations that may appear in recycled material.	37
3.6	Composite visualization of missing stitch defects. Each row corresponds to a defect instance, while columns show the fibular representation, the diffuse channel, and the final annotation overlay.	38
3.7	Annotation protocol difference.	39
3.8	Composite visualization of gap defects. Each row corresponds to a defect instance, while columns show the fibular representation, the azimuthal orientation visualization rendered in HSV, and the final annotation overlay. The annotation distinguishes the gap region from the surrounding fibre distortion induced by the separation between adjacent tows.	40
3.9	Composite visualization of fuzzball defects. Each row corresponds to a defect instance. The fibular representation (left) was used during annotation, while the images on the right show the final polygon overlays.	41
3.10	Composite visualization of drilled tow defects. Each row corresponds to a defect instance. The fibular representation (left) was used during annotation, while the images on the right show the final polygon overlays.	42

3.11	Composite visualization of weaving error defects. Each row corresponds to a defect instance; columns show the fibular representation, the azimuthal channel visualized in HSV, and the final annotation overlay. Weaving errors are primarily identified through local changes in tow spacing and orientation irregularities highlighted in the azimuthal visualization.	43
3.12	Composite visualization of fuzz yarn defects. Each row corresponds to a defect instance. The fibular representation (left) was used during annotation, while the images on the right show the resulting polygon overlays delineating the elongated regions affected by protruding fibres along the yarn direction. .	45
3.13	Composite visualization of distortion defects. Each row corresponds to a defect instance. The fibular representation (left) was used during annotation, while the images on the right show the final polygon overlays marking the region where fibres deviate from the regular textile structure.	46
3.14	Example of a foreign object defect. The fibular channel reveals a localized region whose appearance differs from the surrounding fibre structure. The right image shows the resulting polygon annotation used during dataset construction. The sample also contains a small missing stitch, which is annotated as an additional defect region in the overlay.	47
3.15	Examples of broken tow defects. Each row corresponds to a defect instance. The fibular representation provides a clear visualization of the disrupted fibre bundle, while the overlay images show the final polygon annotations used during dataset construction.	48
4.1	Example of a leakage-prone random split. Samples sharing the same letter correspond to different acquisitions of the same physical defect instance obtained under different positions or orientations. Because correlated acquisitions are distributed across both training and validation subsets, the model may partially memorize defect-specific characteristics during training and later recognize related observations during validation.	52
4.2	Example of a group-consistent split. Samples sharing the same letter correspond to different acquisitions of the same physical defect instance obtained under different positions or orientations. All correlated acquisitions belonging to the same defect group remain confined to a single subset, preventing information leakage across the training-validation boundary.	54

4.3	Conceptual generalization regimes in industrial defect segmentation. Case (A) corresponds to the instance-grouped split evaluated in this work. Case (B) represents generalization to unseen materials and is included as a conceptual future direction. Cases (C) and (D) denote regimes involving previously unseen defect categories.	55
5.1	Comparison of different sensor representations for two defect instances. Columns correspond to the fibular, azimuthal, and diffuse channels, respectively. For visualization purposes, the azimuthal channel is rendered using an HSV colormap to emphasize orientation variations; however, the underlying representation used during training remains the original single-channel orientation tensor.	64
5.2	Example of the frequency spectrum of a carbon-fibre surface obtained through the Fast Fourier Transform.	66
5.3	Examples of frequency-based representations applied to diffuse-channel images.	67
5.4	Visualization of two normalization strategies for the azimuthal channel. Directional normalization preserves angular direction but creates artificial discontinuities at the $-\pi/ + \pi$ boundary. Orientation-only normalization applies a modulo- π mapping that merges opposite directions and yields a smoother representation of fibre orientation.	70
5.5	Visual representation of the CutMix generation process adopted in this work.	80
5.6	Example illustrating the effect of fibre-orientation mismatch during CutMix augmentation.	81
5.7	Example of CutMix augmentation for a missing stitch defect.	82
5.8	Example of physically plausible CutMix-generated composite.	83
6.1	Validation IoU obtained across the ten Monte Carlo seeds for the initial three-channel input configuration under leakage-safe instance-grouped evaluation. Although the architecture and training configuration remain identical, substantial variability emerges depending on the specific train-validation partition generated by each seed. The best-performing seed achieved an IoU of 0.638, whereas the worst-performing seed achieved only 0.412. This behaviour highlights the strong sensitivity of industrial defect segmentation to data partitioning under severe data scarcity and motivates the use of repeated Monte Carlo evaluation instead of single-split performance estimation. . . .	93

6.2	Qualitative comparison of segmentation outputs for different SegFormer backbone capacities. The first column shows the input image with the ground-truth annotation overlay (green). The following columns show the predictions of the B0, B2, and B5 backbones respectively (red).	98
6.3	Qualitative comparison between models trained without and with CutMix augmentation.	100
6.4	Qualitative comparison between YOLO instance segmentation and SegFormer semantic segmentation predictions.	103
6.5	PatchCore anomaly map on a gap defect.	104
6.6	PatchCore anomaly map on a fuzzball defect.	104
6.7	PatchCore anomaly map on a missing stitch defect.	104

List of Tables

3.1	Defect classes present in the dataset and their description.	35
3.2	Number of annotated instances and images per defect class.	35
6.1	Influence of input dimensionality on segmentation performance. Results are reported as mean $\pm$ sample standard deviation across 10 Monte Carlo seeds.	92
6.2	Influence of normalization strategies on segmentation performance. Results are reported as mean $\pm$ sample standard deviation across 10 Monte Carlo seeds.	94
6.3	Influence of frequency-based representations on segmentation performance. Results are reported as mean $\pm$ sample standard deviation across 10 Monte Carlo seeds.	95
6.4	Influence of the azimuthal channel mapping strategy. Results are reported as mean $\pm$ sample standard deviation across 3 Monte Carlo seeds. <i>The numerical values are intended only for relative comparison within this experiment.</i>	96
6.5	Influence of backbone capacity on segmentation performance. Results are reported as mean $\pm$ sample standard deviation across 10 Monte Carlo seeds.	97
6.6	Comparison between the baseline configuration and the same configuration trained with CutMix augmentation.	99
6.7	Comparison between SegFormer and YOLO segmentation models evaluated on the same set of 10 Monte Carlo seeds.	102
6.8	Influence of intermediate fine-tuning on the Bristol Robotics Laboratory dataset before adaptation to the proposed dataset. Results are reported as mean $\pm$ sample standard deviation across 10 Monte Carlo seeds.	106

1 Introduction

Carbon-fibre reinforced composites play an increasingly important role in modern manufacturing due to their high strength-to-weight ratio, corrosion resistance, and favourable mechanical properties. These materials are widely employed in high-performance sectors such as aerospace, automotive, and advanced industrial production, where structural reliability and material quality are critical requirements. The increasing adoption of carbon-fibre composites also raises important challenges related to sustainability, recovery, and reuse at the end of the product lifecycle. In this context, reliable inspection of carbon-fibre materials becomes a key enabling step for circular manufacturing pipelines, where recovered or partially processed material must be evaluated before being reused in subsequent production stages.

However, ensuring the quality of carbon-fibre textile preforms remains a challenging task. Manufacturing processes may introduce several types of defects, including missing stitches, gaps between fibre tows, broken fibres, distortions, or foreign material inclusions. Such defects can compromise downstream processing and negatively affect the structural integrity of the final component. For this reason, the development of reliable inspection methodologies for defect detection and localization represents an important problem from both industrial and research perspectives.

Traditionally, inspection of carbon-fibre materials has relied either on human operators or on handcrafted rule-based image processing systems. Nevertheless, carbon-fibre textiles are visually challenging surfaces. Their appearance depends strongly on fibre orientation, weave pattern, local reflectance behaviour, illumination direction, and material-specific texture. As a consequence, defects are often subtle, heterogeneous, and difficult to distinguish from normal structural variations. These characteristics limit the robustness of manual inspection and make the design of classical handcrafted inspection pipelines particularly difficult.

In recent years, deep learning approaches have shown strong potential for automated visual inspection in industrial settings. Compared to rule-based systems, learning-based models can capture more complex visual patterns and potentially adapt better to the variability of real manufacturing conditions. However, the direct application of deep learning to carbon-fibre inspection is far from straightforward. Unlike standard natural-image computer vision prob-

lems, industrial inspection typically operates under severe data constraints, specialized sensing modalities, and acquisition procedures that may introduce hidden correlations within the dataset.

1.1 Problem Statement

In the specific context of this thesis, the image data is acquired using the PROFACTOR Fscan inspection system, which produces multiple sensor-derived channels designed for carbon-fibre material inspection rather than conventional RGB imagery.

One of the main challenges encountered during the project was the limited availability of independent industrial defect samples. From the perspective of supervised deep learning, the dataset initially appeared extremely small relative to the complexity of the target segmentation task, particularly considering the high variability of defect appearance and the specialized nature of the sensor representations.

At the same time, previously reported classification results on a related industrial dataset suggested that learning-based approaches could still achieve strong performance despite severe data limitations. These observations initially motivated the development of supervised semantic segmentation models under the assumption that the available data could support practical defect localization.

However, subsequent analysis revealed that the industrial acquisition protocol introduced significant correlations between samples, which could substantially influence performance evaluation if not handled carefully. As a consequence, the work evolved from a straightforward segmentation study into a broader methodological investigation focused not only on defect segmentation itself, but also on reliable evaluation and generalization under realistic industrial conditions.

1.2 From Classification to Segmentation

The starting point of this work was an existing industrial dataset originally associated with a classification-oriented study. In that context, the primary objective was defect classification, namely predicting whether a given sample contained a specific defect category. Although segmentation annotations were already available, they had been introduced primarily as supporting information for the classification pipeline rather than as the target of a dedicated semantic segmentation task. In particular, the segmentation masks were used to provide spatial information about the defective regions while the final learning objective remained image-level defect classification.

From an industrial perspective, however, image-level classification provides only limited information. Beyond simple defect presence detection, pixel-level localization can provide important operational advantages in industrial inspection workflows.

In practical manufacturing environments, it is often necessary not only to determine whether a material sample is defective, but also to estimate the spatial extent, geometry, and precise location of the defective region. Such information can support several downstream operations. For example, localized defect masks may be used to guide manual inspection, estimate defect severity, identify regions requiring material rejection or repair, support traceability and quality documentation, or enable automated downstream processing decisions. In robotic or inline inspection settings, spatial defect localization may also facilitate targeted intervention strategies, such as removal of defective regions or adaptive process control.

For these reasons, semantic segmentation provides a substantially richer and more operationally useful representation than image-level classification alone, particularly in industrial inspection scenarios where the geometry and extent of the defect are important aspects of the decision-making process. Unlike classification approaches, semantic segmentation requires the model not only to recognize the presence of a defect but also to infer its precise spatial structure at pixel level. This formulation introduces substantially greater complexity. In particular, segmentation is inherently more sensitive to annotation quality, foreground-background imbalance, and dataset variability than image-level classification. These challenges become especially critical in industrial settings where the number of independent defect instances is limited and where annotations may exhibit inconsistencies across acquisitions. The transition from classification to segmentation therefore required reconsidering several aspects of the original pipeline, including:

- the suitability and consistency of the available annotations;
- the structure and acquisition protocol of the dataset;
- the representation of the sensor-derived channels;
- the evaluation methodology used to assess model generalization;
- and the robustness of learning strategies under severe data scarcity.

1.3 Research Objectives

Motivated by the industrial and methodological considerations introduced above, this thesis pursues a set of interconnected research objectives.

The first objective is to improve the suitability of the available industrial data for supervised learning through dataset curation, controlled acquisition procedures, and annotation refinement. In particular, the thesis seeks to move beyond the limitations of the original classification-oriented dataset by constructing a new dataset specifically designed to support supervised semantic segmentation.

The second objective is to establish a reliable evaluation methodology for industrial defect inspection under grouped acquisition conditions. In particular, the work investigates how correlated acquisitions can introduce information leakage between training and validation data and studies splitting strategies designed to provide realistic estimates of model generalization.

The third objective is to study how the sensor-derived channels produced by the Fscan system should be represented and normalized when used as input to deep learning models. Since the considered channels encode different physical properties of the material surface, the thesis investigates alternative channel combinations and normalization strategies in order to determine which representations are most informative for defect localization.

The fourth objective is to investigate the feasibility of supervised semantic segmentation for carbon-fibre defect inspection under severe industrial data scarcity. Rather than focusing exclusively on maximizing segmentation accuracy, the thesis aims to understand which methodological choices are necessary to obtain reliable and practically meaningful performance in constrained industrial settings.

The fifth objective is to investigate the suitability of modern segmentation architectures and training strategies for this industrial setting. This includes studying the influence of backbone capacity, optimization choices, imbalance-aware loss formulations, and data augmentation strategies under realistic small-data conditions.

Finally, the thesis aims to provide a systematic experimental analysis of the principal factors affecting segmentation performance in industrial carbon-fibre inspection, with the broader goal of identifying robust methodological directions for reliable learning-based inspection systems.

1.4 Main Contributions

The work presented in this thesis provides the following main contributions.

- **Industrial dataset acquisition and curation.** Starting from an existing industrial dataset that was not fully suitable for dedicated supervised segmentation, a new dataset of carbon-fibre textile defects was acquired and curated using a controlled inspection protocol. The resulting dataset includes multiple acquisitions of defect instances together with pixel-level annotations and provides a substantially more suitable basis for supervised segmentation studies.
- **Leakage-safe evaluation methodology.** Industrial inspection datasets often contain multiple acquisitions of the same physical defect captured under different positions or orientations. If such correlated samples are randomly distributed across training and validation sets, performance estimates can become artificially inflated. This thesis experimentally investigates this phenomenon and introduces leakage-safe splitting protocols designed to provide more realistic estimates of model generalization.
- **Analysis of industrial sensor representations.** Different combinations of sensor-derived channels and normalization strategies are investigated in order to study how physically meaningful representations such as fibre orientation and reflectance information can be effectively integrated into deep-learning-based segmentation pipelines.
- **Systematic analysis of segmentation methodology under industrial data scarcity.** The thesis presents a broad experimental investigation of architecture selection, backbone capacity, optimization strategies, and imbalance-aware loss formulations under realistic small-data industrial conditions.
- **Adaptation of CutMix augmentation for industrial defect segmentation.** A CutMix-based augmentation strategy is adapted to the industrial inspection setting in order to increase training diversity and improve robustness under limited supervision conditions.

Taken together, these contributions provide a structured investigation of how supervised deep learning can be adapted to the challenging setting of carbon-fibre defect segmentation, where the primary difficulties arise not only from model selection but also from data representation, annotation quality, acquisition protocols, and evaluation methodology.

1.5 Thesis Organization

The remainder of this thesis is organized as follows.

Chapter 2 presents the theoretical background and related work relevant to industrial visual inspection, semantic segmentation, and industrial defect detection paradigms. Particular attention is given to modern segmentation architectures, industrial inspection constraints, and methodological considerations related to small-data learning and correlated industrial acquisitions.

Chapter 3 describes the industrial acquisition setting and analyzes the characteristics and limitations of the original industrial dataset. It presents the sensing system, acquisition protocol, annotation methodology, dataset organization, and qualitative analysis of the considered defect categories, together with the dataset design choices motivated by the identified annotation and acquisition issues.

Chapter 4 focuses on the problem of evaluation reliability under grouped industrial acquisitions. Different dataset splitting strategies are analysed, and the impact of correlated acquisitions on performance estimation is experimentally investigated through a reproduction study of the original classification setting.

Chapter 5 introduces the experimental and methodological framework adopted for supervised semantic segmentation. The chapter discusses the selected sensor representations, normalization strategies, segmentation architecture, optimization choices, augmentation techniques, evaluation metrics, and alternative approaches explored during the study.

Chapter 6 presents the experimental results and analyses the influence of representation design, architecture configuration, optimization choices, and augmentation strategies on segmentation performance. The chapter also discusses exploratory experiments including Patch-Core anomaly detection, YOLO-based segmentation, and cross-dataset transfer learning.

Finally, Chapter 7 summarizes the main findings of the thesis, discusses the principal limitations of the current work, and outlines possible directions for future research.

2 Background and Related Work

This chapter presents the theoretical background and related research needed to frame the defect segmentation problem addressed in this thesis. It first discusses the role of automated visual inspection in industrial environments and then reviews the main defect detection paradigms used in computer vision, namely image classification, object detection, segmentation, and anomaly-based approaches. The chapter also introduces the segmentation model families most relevant to this work and summarizes the principal challenges that characterize industrial defect inspection, with particular focus on texture-rich materials and limited-data settings. Finally, it reviews previous studies on textile and carbon-fibre defect analysis in order to position the present work within the existing literature.

2.1 Automated visual inspection in industrial settings

Visual inspection is a fundamental component of industrial quality assurance. In many manufacturing and material-processing scenarios, the ability to identify defects at an early stage is essential to ensure product reliability, reduce waste, and prevent defective items from propagating through subsequent processing steps. Modern industrial inspection systems increasingly rely on computer-vision and machine-learning techniques in order to automate defect detection and improve inspection consistency.

Depending on the application domain, inspection pipelines may operate on standard RGB imagery or on data acquired through specialized sensing systems designed to capture material-specific physical properties. Compared with mainstream natural-image computer vision, industrial inspection often presents additional challenges related to limited annotated datasets, highly structured textures, subtle defect appearance, and domain-specific acquisition protocols. Furthermore, the required level of output detail may vary substantially depending on the industrial application, ranging from simple image-level defect classification to precise pixel-wise localization of defective regions.

For these reasons, industrial defect inspection can be formulated through different computer-vision paradigms, each characterized by different annotation requirements, output

representations, and levels of spatial precision. Before discussing the specific methodology adopted in this thesis, it is therefore useful to review the main defect detection paradigms commonly employed in computer vision and industrial inspection.

2.2 Defect detection paradigms

Defect detection can be formulated at different levels of granularity depending on the required output, the available annotations, and the intended industrial use case. In computer vision, four major paradigms are particularly relevant in this context: image classification, object detection, semantic segmentation, and anomaly-based approaches. Each of these paradigms makes different assumptions about the data and provides a different level of spatial information.

2.2.1 Image classification

Image classification is the simplest formulation of visual defect detection. Given an input image, the model predicts one or more labels describing the image as a whole. In the most basic defect inspection setting, this corresponds to a binary decision indicating whether the image is defective or non-defective. More elaborate variants may predict the specific defect category at image level. The main advantage of image classification lies in its relatively simple supervision requirements. Compared with more detailed paradigms, image-level labels are cheaper and faster to obtain than bounding boxes or pixel-wise annotations. This makes classification attractive in industrial settings where data annotation is expensive or where an initial feasibility study is needed. Classification models are also comparatively simple to train and evaluate, and can be effective when the objective is only to trigger an alarm or reject an entire sample. However, classification provides no explicit information about *where* the defect is located inside the image. This limitation becomes particularly relevant when defects occupy only a small fraction of the image, when multiple defects may be present simultaneously, or when the spatial extent of the anomaly is important for downstream decisions. In such cases, image-level predictions are often insufficient, since they do not indicate the position, size, or shape of the defective region. For this reason, although classification can be a useful first step, it is generally too coarse for applications requiring precise localization of industrial defects.

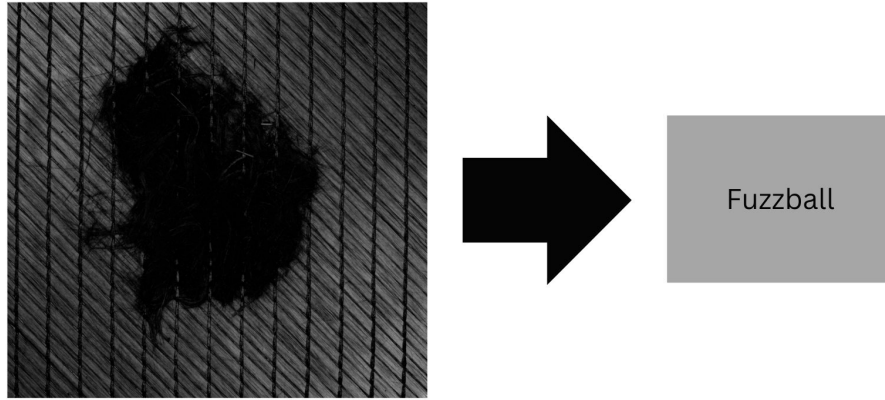


Figure 2.1: Conceptual illustration of image classification for industrial defect inspection. The input image is assigned a single defect label at image level, while no explicit localization of the defective region is provided.

2.2.2 Object detection

Object detection extends image classification by introducing coarse spatial localization. Instead of assigning a single label to the whole image, the model predicts one or more bounding boxes together with the associated class labels. In industrial inspection, this allows the system not only to determine whether a defect is present, but also to indicate approximately where it appears. This paradigm is well suited to problems in which defective regions can be reasonably approximated by compact objects with separable spatial extent. Object detection can be particularly useful when the goal is to count defect instances, identify their approximate location, or support rapid screening in real-time environments. In comparison with semantic segmentation, bounding-box annotation is typically less expensive and less time-consuming to produce, while still providing a stronger localization signal than simple image-level labels. Nevertheless, object detection has important limitations in defect inspection scenarios involving elongated, irregular, or diffuse anomalies. Many industrial defects do not exhibit well-defined object boundaries and may instead appear as thin structural irregularities, subtle surface distortions, or fragmented regions embedded in a textured background. In such cases, enclosing the anomaly in a rectangular box may be an imprecise representation, particularly when the exact spatial extent of the defect matters. As a result, object detection can provide useful coarse localization, but may not be sufficiently expressive for tasks requiring detailed pixel-level delineation.

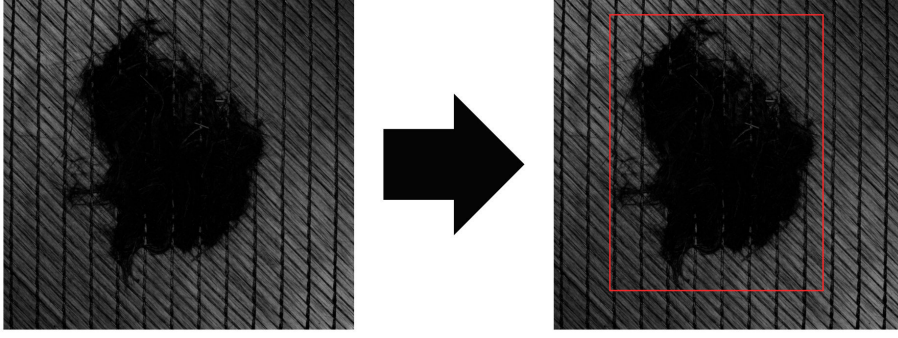


Figure 2.2: Conceptual illustration of object detection for industrial defect inspection. The model predicts a bounding box around the defective region, providing approximate spatial localization without explicitly delineating the precise defect boundaries.

2.2.3 Segmentation-based approaches

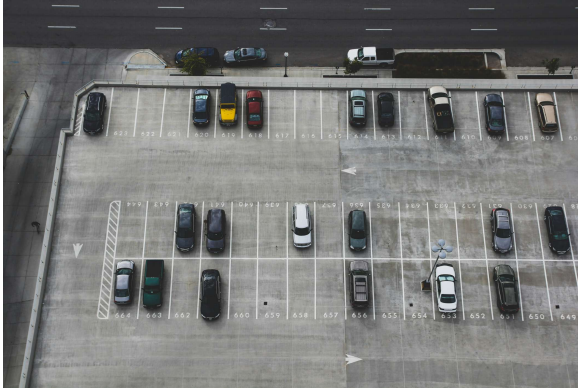
Segmentation tasks aim to assign a label to image regions at pixel level, thus providing a much finer spatial description than image classification or object detection. Depending on the type of output required, segmentation approaches are commonly divided into three main categories: *semantic segmentation*, *instance segmentation*, and *panoptic segmentation*.

In **semantic segmentation**, each pixel is assigned to a semantic class, but different instances of the same class are not distinguished from one another. For example, if multiple defects are present in the same image, semantic segmentation assigns a class label to each pixel but does not explicitly distinguish between different defect instances. Consequently, multiple defects belonging to the same class are represented as a single connected prediction mask whenever their regions touch or overlap spatially.

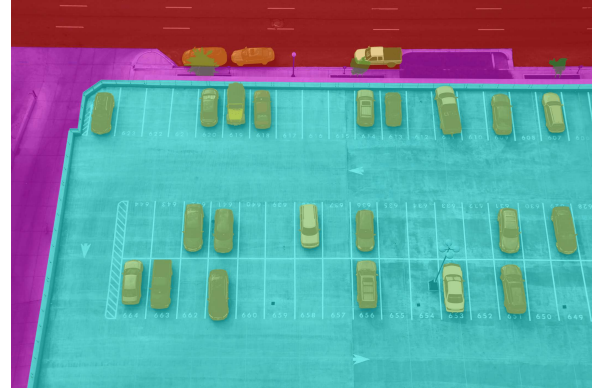
In **instance segmentation**, the goal is not only to classify pixels, but also to distinguish different object instances belonging to the same class. Thus, two spatially separate defects of the same category would be assigned two different masks, even though they share the same semantic label. This formulation is particularly useful when the identification, counting, or independent tracking of individual instances is required.

Panoptic segmentation combines the two previous paradigms into a unified framework. Each pixel receives a semantic label, as in semantic segmentation, while countable object-like regions are also separated into individual instances, as in instance segmentation. In other words, panoptic segmentation provides both complete scene labeling and instance-level differentiation where needed.

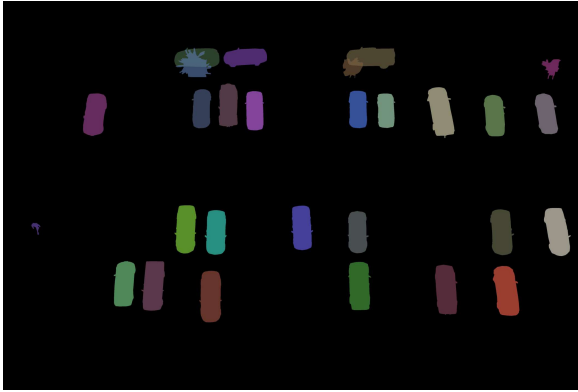
In the present thesis, the problem is formulated as a **binary semantic segmentation** task, where each pixel is classified as either *defect* or *background*. Although the annotation process preserved information about the defect categories at polygon level, all defect types were collapsed into a single foreground class for model training and evaluation. This choice was moti-



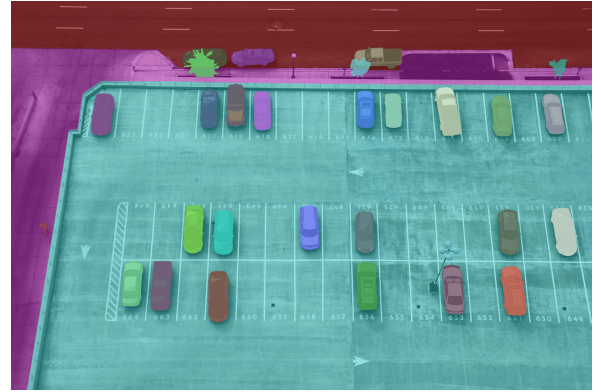
(a) Input image



(b) Semantic segmentation



(c) Instance segmentation



(d) Panoptic segmentation

Figure 2.3: Conceptual comparison between the main segmentation paradigms. Semantic segmentation assigns a class label to every pixel in the image, without distinguishing different instances of the same class. Instance segmentation identifies and separates individual countable object instances, while background or uncountable regions are not explicitly labeled in this visualization. Panoptic segmentation combines both perspectives by assigning a semantic label to every pixel and distinguishing individual instances for countable objects.

vated by both practical and methodological considerations.

First, the primary objective of the work is to localize defective regions accurately, rather than to count individual instances or to distinguish adjacent defects of the same type. Therefore, semantic segmentation provides the level of output detail that is most relevant for the intended inspection task. Second, the available dataset is relatively small, and the addition of instance-level or defect-type discrimination would substantially increase the complexity of the learning problem. Requiring the model not only to detect defective regions, but also to distinguish between multiple defect categories or separate individual instances, would substantially increase the complexity of the learning problem and would likely require larger annotated datasets and stronger supervision in order to achieve reliable generalization. Under limited-data conditions, collapsing all defects into a single foreground class makes the task more tractable and allows the model to focus on learning the distinction that is most important in this setting, namely whether a pixel belongs to a defective region or not. For these reasons, semantic segmentation represents the most appropriate formulation for the present work, offering a suitable balance between spatial precision, annotation requirements, and learning complexity.

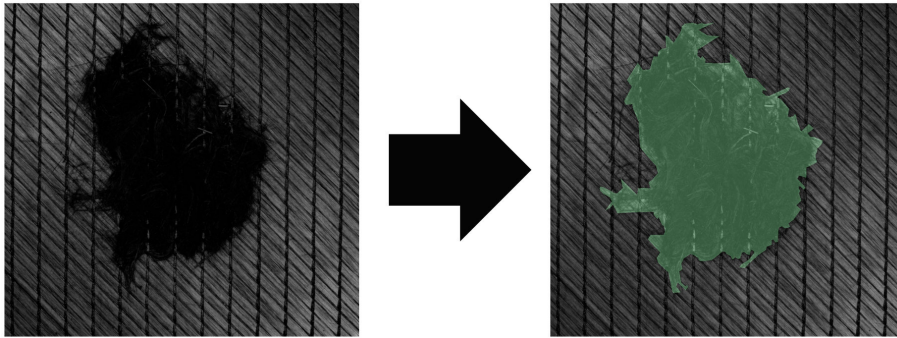


Figure 2.4: Illustration of semantic segmentation. Unlike image classification and object detection, the model produces a pixel-level prediction mask that explicitly captures the shape and spatial extent of the defect.

2.2.4 Anomaly detection and self-supervised approaches

Anomaly detection addresses defect inspection from a different perspective. Rather than learning to predict explicit defect labels from fully annotated examples, anomaly-based methods typically model the appearance of normal data and identify deviations from this learned normality as potential defects. In industrial inspection, this paradigm is attractive because defective samples are often scarce, heterogeneous, and costly to annotate, whereas normal samples are generally easier to collect in larger quantity.

Classical anomaly detection methods relied on handcrafted descriptors and statistical modelling, while more recent approaches increasingly employ deep features, reconstruction objectives, feature-space distance measures, or self-supervised representation learning. In this broader context, self-supervised approaches aim to learn useful visual representations from unlabeled data by solving auxiliary pretext tasks or by enforcing consistency across transformed views of the same input. These learned representations can then be exploited for anomaly scoring or adapted to downstream tasks with limited supervision.

The main advantage of anomaly-based and self-supervised methods is their reduced dependence on dense labeled defect data. This makes them particularly appealing in industrial scenarios where annotation is expensive or where only a few examples of rare defects are available. However, these approaches also have limitations. Their performance may depend strongly on how well the normal data distribution is represented, and subtle structural variations in highly textured materials can make the distinction between normality and anomaly ambiguous. In addition, when the final goal is precise delineation of defective regions, anomaly maps may be less accurate or less stable than predictions obtained from fully supervised segmentation models.

For these reasons, anomaly detection and self-supervised learning represent important alternative paradigms, especially under severe annotation constraints. Nevertheless, when reliable pixel-level annotations are available and the objective is accurate spatial localization, supervised semantic segmentation generally offers a more direct and interpretable formulation of the problem.

2.3 Semantic segmentation models relevant to this work

A wide range of architectures has been proposed for semantic segmentation, spanning both convolutional and Transformer-based approaches. In the context of this thesis, the objective is not to provide an exhaustive survey of all existing models, but rather to introduce the segmentation architectures most closely related to the methodological choices explored in this work. In particular, this section focuses primarily on SegFormer [1], which constitutes the main semantic segmentation architecture adopted throughout the experimental study, and more briefly on YOLO-based segmentation variants [2], which were investigated as an alternative instance-segmentation approach. These model families differ substantially in how they represent spatial information and generate segmentation outputs. SegFormer directly predicts dense pixel-wise probability maps over the entire image, whereas YOLO-based segmentation architectures first detect object instances and subsequently generate object-specific segmentation masks. Such differences are especially relevant in industrial defect inspection scenarios, where defects may

be small, irregular, and embedded within highly structured carbon-fibre textures. For this reason, understanding the main design principles behind these architectures is useful in order to motivate the methodological choices and experimental analyses presented in the remainder of the thesis.

2.3.1 SegFormer

SegFormer is a Transformer-based semantic segmentation architecture proposed as a simple and efficient alternative to earlier Vision Transformer segmentation models. Its design combines a hierarchical Transformer encoder with a lightweight all-MLP decoder, allowing the network to capture both local structure and global context while keeping the decoder computationally simple.

Compared with earlier Transformer-based segmentation approaches, SegFormer is characterized by three particularly important design choices:

- a **hierarchical multi-stage encoder**, which produces feature maps at different spatial resolutions similarly to convolutional backbones;
- the absence of **explicit positional embeddings**, making the architecture more flexible with respect to input resolution;
- a **lightweight all-MLP decoder**, which fuses multi-scale features without relying on heavy convolutional modules such as ASPP or complex feature pyramid decoders.

These properties make SegFormer particularly attractive in dense prediction problems where both global context and fine spatial detail are important, while computational efficiency and transfer learning remain relevant.

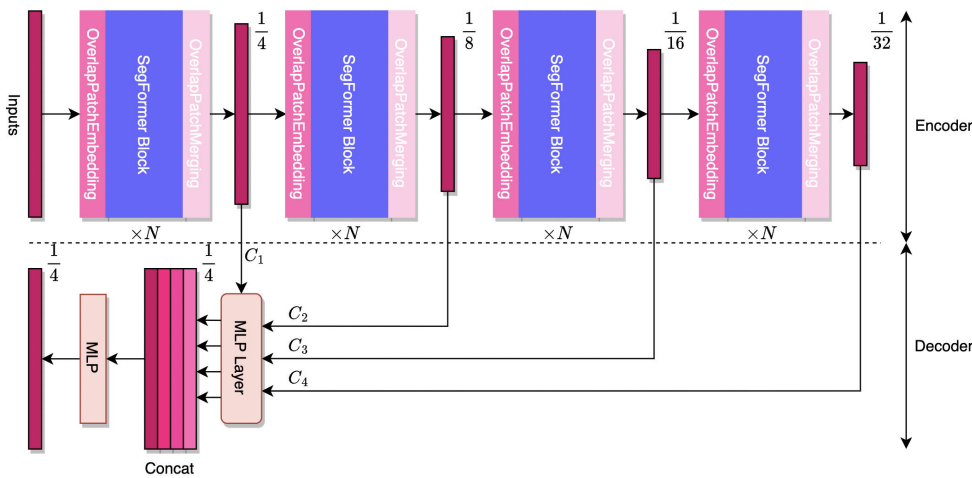


Figure 2.5: SegFormer encoder–decoder architecture.

Overall architecture

Let the input image be denoted by

$$I \in \mathbb{R}^{H \times W \times 3}.$$

The SegFormer architecture is composed of:

1. a **hierarchical Transformer encoder**, which extracts four feature maps at progressively lower spatial resolutions;
2. a **lightweight all-MLP decoder**, which projects, upsamples, concatenates, and fuses these feature maps in order to predict the final segmentation mask.

The encoder outputs a set of multi-scale representations

$$\{F_1, F_2, F_3, F_4\},$$

where the spatial resolution decreases from stage to stage, while the channel dimension increases. This organization is analogous in spirit to convolutional backbones used in semantic segmentation, but it is implemented through Transformer blocks.

Hierarchical Transformer encoder

A key difference between SegFormer and earlier Vision Transformer models is that the encoder is *hierarchical*. Instead of processing all image tokens at a single fixed resolution, the encoder is organized into four stages. At each stage, the spatial resolution is reduced and the feature dimension is increased, allowing the network to build progressively more abstract representations.

Overlapped patch embedding. Unlike the original Vision Transformer (ViT), which partitions the image into non-overlapping patches, SegFormer employs *overlapped patch embedding* in the first encoder stage implemented as a convolution:

$$F_1 = \text{Conv}_{7 \times 7}^{S=4}(I),$$

where the kernel size is 7×7 , the stride is 4, and padding is applied appropriately.

This operation produces a token grid of spatial size

$$\frac{H}{4} \times \frac{W}{4},$$

with an initial channel dimension of 64.

Although SegFormer is often described as using a patch size of 4×4 , this should be interpreted as the effective token spacing induced by the stride, rather than as the actual receptive field of each token. Because the convolution kernel is larger than the stride, neighbouring tokens are generated from overlapping image regions.

This overlap is important for two reasons:

- it preserves local continuity between adjacent tokens;
- it reduces the need for explicit positional encoding, since the local image structure is already embedded into the tokenization process.

Multi-stage structure. After the first stage, the encoder continues with three additional stages. Between consecutive stages, a further convolutional downsampling step, referred to as *overlapped patch merging*, halves the spatial resolution and increases the feature dimension. The encoder therefore produces a hierarchy of feature maps:

$$F_1 \in \mathbb{R}^{\frac{H}{4} \times \frac{W}{4} \times 64}, \quad F_2 \in \mathbb{R}^{\frac{H}{8} \times \frac{W}{8} \times 128}, \quad F_3 \in \mathbb{R}^{\frac{H}{16} \times \frac{W}{16} \times 320}, \quad F_4 \in \mathbb{R}^{\frac{H}{32} \times \frac{W}{32} \times 512}.$$

SegFormer block

Each stage of the encoder consists of a stack of SegFormer blocks. Each block contains two main components:

- **Efficient Multi-Head Self-Attention;**
- **Mix-FFN.**

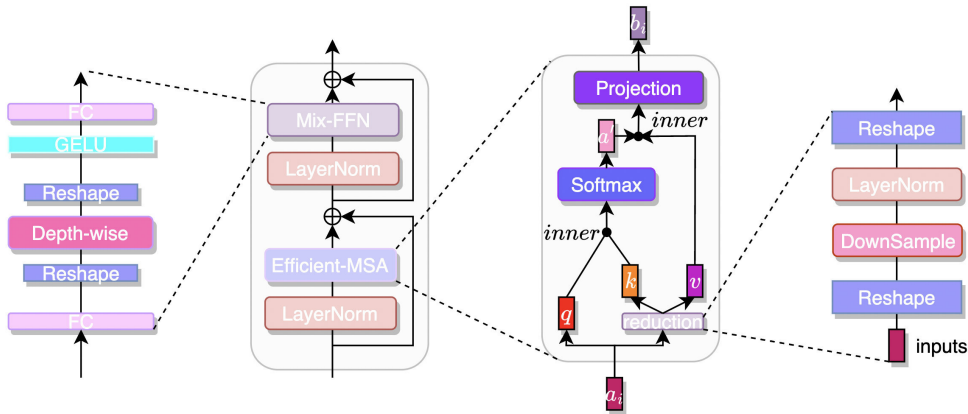


Figure 2.6: Structure of a SegFormer block.

Efficient Multi-Head Self-Attention. In standard self-attention, the computational complexity grows quadratically with the number of tokens $\mathcal{O}(N^2)$, where N is the token count. This becomes computationally expensive for dense prediction tasks, especially at high spatial resolutions.

SegFormer reduces this cost by applying a spatial reduction to the feature maps used to generate the key and value sequences before computing attention, while the query sequence remains at full resolution. In practice, this corresponds to spatially downsampling the key and value representations, thereby reducing the number of tokens participating in the attention computation. For example, consider an early encoder stage containing $64 \times 64 = 4096$ tokens. In standard self-attention, each query token attends to all other tokens, producing an attention matrix of size 4096×4096 , which is computationally expensive. If a spatial reduction ratio $R = 64$ is applied, the key and value sequences are reduced from 4096 tokens to $\frac{4096}{64} = 64$ tokens, while the query sequence still contains all 4096 tokens. The resulting attention matrix therefore becomes 4096×64 instead of 4096×4096 .

As a consequence, the computational complexity is reduced approximately to

$$\mathcal{O}\left(\frac{N^2}{R}\right).$$

Different encoder stages use different reduction ratios, typically $R \in \{64, 16, 4, 1\}$, so that the early high-resolution stages are made substantially cheaper, while later low-resolution stages can afford more complete attention. This mechanism allows SegFormer to preserve the global-context modeling capability of self-attention while making the architecture computationally practical for semantic segmentation.

Why SegFormer does not use explicit positional embeddings

A particularly innovative aspect of SegFormer is that it avoids explicit positional embeddings. In standard Transformer architectures, tokens are processed as a set, and self-attention alone is permutation-invariant. For this reason, models such as ViT require explicit positional encodings to inform the network about spatial order. SegFormer reduces this dependence through two architectural mechanisms.

First source of spatial information: overlapped patch embedding. Because tokens are extracted through overlapping convolutions, each token already reflects a local image region together with its neighbourhood. This introduces local spatial bias directly during tokenization.

Second source of spatial information: Mix-FFN. In vanilla Transformers, each token is passed through a feed-forward network of the form

$$x_{\text{out}} = \text{MLP}(\text{GELU}(\text{MLP}(x))) + x,$$

where the two MLP layers correspond to linear channel projections applied independently to each token. Since this operation is performed separately for every token, it mixes only channel dimensions and does not model local spatial relations between neighbouring tokens.

SegFormer replaces this with a **Mix-FFN**, defined as

$$x_{\text{out}} = \text{MLP}(\text{GELU}(\text{DWConv}_{3 \times 3}(\text{MLP}(x)))) + x.$$

The inserted depthwise 3×3 convolution introduces local interaction between neighbouring tokens directly inside the feed-forward path, thereby injecting spatial structure into the Transformer block and making the encoder more suitable for dense prediction tasks such as semantic segmentation. Moreover, because the convolution uses padding at image boundaries, tokens near the border experience a different effective context from tokens in the interior. This breaks perfect spatial symmetry and allows the network to infer positional cues implicitly, a phenomenon often described as positional information *leaking through zero padding*. In this context, this is not a defect but a useful source of spatial awareness. As a result, SegFormer does not require explicit positional embeddings to preserve spatial structure and avoids the interpolation issues that can arise when positional encodings must be adapted to different resolutions at test time.

Lightweight all-MLP decoder

Unlike CNN-based decoders such as ASPP or UPerNet, which rely on more complex convolutional modules, SegFormer adopts a lightweight all-MLP decoder. The decoder operates as follows:

- each encoder feature map F_i is passed through a linear layer in order to unify the channel dimension;
- the resulting feature maps are upsampled to a common spatial resolution, namely $\frac{H}{4} \times \frac{W}{4}$;
- the upsampled features are concatenated;
- a further linear layer is used to fuse the concatenated representation;
- a final linear layer predicts the segmentation logits.

Formally, the decoder can be written as

$$\begin{aligned}\hat{F}_i &= \text{Linear}(C_i, C)(F_i), \quad i = 1, \dots, 4, \\ \hat{F}_i^\uparrow &= \text{Upsample}\left(\frac{H}{4} \times \frac{W}{4}\right)(\hat{F}_i), \\ F &= \text{Linear}(4C, C)\left(\text{Concat}(\hat{F}_1^\uparrow, \hat{F}_2^\uparrow, \hat{F}_3^\uparrow, \hat{F}_4^\uparrow)\right), \\ M &= \text{Linear}(C, N_{\text{cls}})(F),\end{aligned}$$

where N_{cls} denotes the number of segmentation classes.

The decoder output:

$$M \in \mathbb{R}^{\frac{H}{4} \times \frac{W}{4} \times N_{\text{cls}}}.$$

is then upsampled by bilinear interpolation to the original spatial resolution (H, W) .

This design is much simpler than heavy convolutional decoders, and relies on the fact that the hierarchical Transformer encoder already provides strong multi-scale contextual features.

Why SegFormer is relevant to this thesis

SegFormer is particularly relevant for the problem addressed in this thesis for several reasons. First, industrial defect segmentation requires distinguishing subtle anomalous regions from highly structured background textures. This benefits from a model capable of combining local structural information with broad contextual reasoning.

Second, the architecture is well suited to transfer learning. Since the available industrial dataset is relatively limited, pretrained variants of SegFormer offer a practical advantage over training large segmentation models from scratch.

Third, the hierarchical nature of the encoder makes the architecture appropriate for defects that vary strongly in scale, from very small localized anomalies to larger structural distortions. Finally, the lightweight decoder and the efficient attention design make SegFormer computationally attractive while preserving strong dense prediction performance. For these reasons, SegFormer was selected as the main semantic segmentation architecture investigated in this thesis.

2.4 Challenges in industrial defect inspection

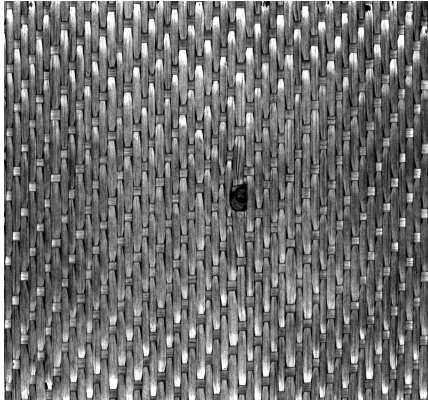
Industrial defect inspection presents a number of challenges that distinguish it from mainstream computer vision settings based on large-scale natural-image datasets. In many industrial applications, defects are rare, subtle, and highly variable in appearance, while the available data

is often limited, costly to annotate, and acquired through specialized sensing systems rather than standard RGB cameras. As a result, the design and evaluation of learning-based inspection methods must take into account not only model architecture, but also the statistical and operational constraints imposed by the data itself. These difficulties are particularly relevant in the context of texture-rich materials, where normal appearance may already exhibit strong local variation and where the visual distinction between defective and non-defective regions may be extremely subtle. In such settings, the learning problem is shaped by a combination of data imbalance, scarce supervision, sensor-specific representations, and acquisition protocols that may unintentionally introduce hidden correlations.

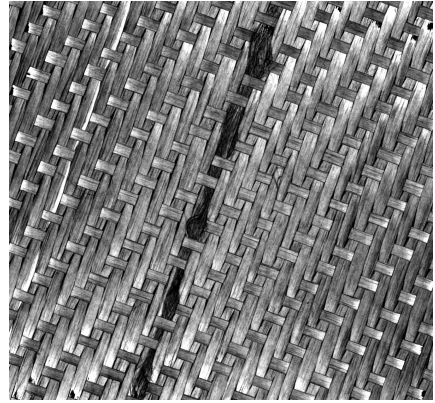
The following sections summarize the main challenges that are especially relevant for the problem addressed in this thesis.

2.4.1 Class imbalance and small defect regions

A common characteristic of industrial defect inspection is the strong imbalance between normal and defective regions. In many datasets, the overwhelming majority of pixels correspond to background or defect-free material, whereas defects occupy only a small fraction of the image. This imbalance is particularly severe in segmentation tasks, where the target class may consist of very small localized structures surrounded by large areas of normal texture. Such a distribution creates difficulties both during training and during evaluation. From the optimization perspective, a model trained on highly imbalanced data may be biased toward the dominant background class, since predicting background over most pixels already yields a deceptively high apparent accuracy. As a consequence, standard loss formulations may fail to sufficiently emphasize the defective regions, especially when these are small or sparse. From the evaluation perspective, metrics that do not account properly for class imbalance may produce an overly optimistic impression of performance. The challenge becomes even more pronounced when defects vary significantly in spatial extent, as illustrated in Figure 2.7.



(a) Example of a small localized defect.



(b) Example of an elongated defect.

Figure 2.7: Examples of defect regions with different spatial extent. Industrial defects may range from very small localized anomalies to larger elongated structures, making robust segmentation more challenging.

2.4.2 Limited annotated datasets

Another major challenge in industrial inspection is the limited availability of annotated data. Unlike large-scale natural-image benchmarks, industrial datasets are rarely collected at massive scale, since data acquisition typically depends on specialized equipment, controlled inspection setups, and expert intervention. In addition, the annotation process is often expensive and time-consuming, particularly for dense tasks such as semantic segmentation, where precise delineation of defective regions may require domain knowledge and careful manual inspection. This scarcity increases the risk of overfitting, reduces the diversity of observable defect appearances, and makes performance estimates more sensitive to the specific train-validation split.

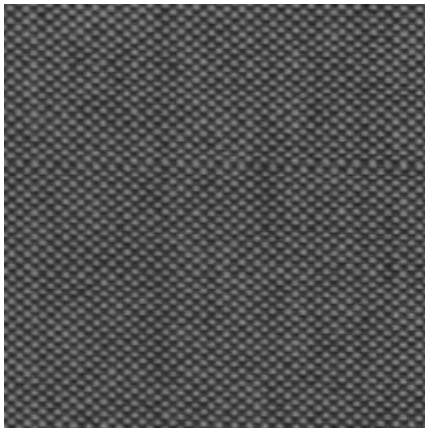
The problem is further compounded by the heterogeneity of industrial defects, which may differ substantially in shape, contrast, texture, and visibility even within the same general class. As a result, limited datasets may fail to capture the full variability expected during deployment. For this reason, transfer learning, suitable augmentation strategies, and carefully designed validation protocols become especially important in industrial defect inspection.

2.4.3 Domain shift and non-standard sensor modalities

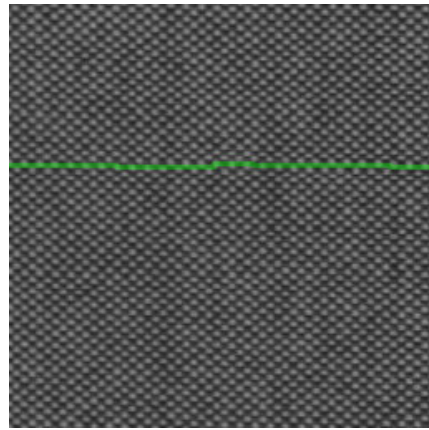
Industrial vision systems frequently operate under conditions that differ substantially from those assumed in mainstream computer vision benchmarks. In many cases, the input data is not composed of conventional RGB images, but of measurements acquired through specialized sensors designed to capture material-specific reflectance, geometry, or other physical properties. This introduces a form of domain mismatch with respect to the large natural-image

datasets on which many modern deep learning models are pretrained.

From a representation-learning perspective, this mismatch can be significant. Models pretrained on natural images implicitly encode statistical regularities. Industrial sensor data, by contrast, may exhibit very different low-level statistics and may encode information in modalities that do not have a direct analogue in natural-image datasets. As a consequence, transfer learning remains useful but cannot be assumed to be fully aligned with the target domain. A related issue is that the visual appearance of the inspected material may vary not only because of defect type, but also because of changes in acquisition conditions, surface orientation, illumination geometry, or intrinsic material properties. In texture-rich inspection scenarios, this may make the distinction between normal and defective regions particularly difficult, since normal samples themselves can exhibit strong local variation. This difficulty is illustrated in Figure 2.8, where the defect is visually subtle and mainly corresponds to a slight structural irregularity in the textile pattern rather than to an obvious change in color or contrast.



(a) A textile sample from the AITEX dataset containing a subtle defect.



(b) The same sample with the ground-truth defect region highlighted.

Figure 2.8: Example of a subtle textile defect from the AITEX dataset [3]. In this case, the anomaly is difficult to identify visually because it mainly appears as a local variation in the spacing or structure of the textile pattern, rather than as a strong change in color or overall appearance. The right image highlights the annotated defect region through the ground-truth mask.

Therefore, the challenge is not simply to recognize a defect class in isolation, but to distinguish subtle local irregularities from the structured variability of the background material. In the context of this thesis, these considerations are especially important because the data is acquired through a specialized industrial sensing system that provides multiple derived channels rather than standard RGB inputs. This makes the design of a suitable input representation a central part of the problem, since the choice of channels, preprocessing strategy, and normal-

ization scheme may strongly affect the ability of the model to separate defect patterns from the surrounding material structure.

2.4.4 Data leakage and correlated acquisitions

Data leakage occurs when information that should not be available during model training indirectly influences the training process or the validation procedure. In supervised learning, this can lead to overly optimistic performance estimates, because the validation data is no longer statistically independent from the information used to fit or select the model.

A particularly relevant form of leakage arises when samples treated as independent actually share hidden dependencies. This situation is common in industrial inspection, where the acquisition protocol itself may introduce correlations between samples. For example, the same physical defect instance can be captured multiple times under different positions, orientations, or sensor configurations. Although such samples may appear as distinct images at first glance, they may still contain highly overlapping information.

If these correlated acquisitions are randomly distributed across training and validation sets, the resulting evaluation may become biased. More specifically, the model may achieve artificially high validation performance not because it has learned to generalize to unseen defect instances, but because it has already observed very similar views of the same underlying physical anomaly during training. In such a case, the validation split no longer reflects the true difficulty of generalization and may lead to overly optimistic conclusions about model effectiveness.

The issue is particularly relevant in small industrial datasets, where repeated acquisitions can represent a large fraction of the available data. In this setting, defining the train–validation split is not merely a procedural detail, but an essential part of the experimental methodology. Preventing this form of leakage requires accounting for the grouped structure of the acquisitions, so that correlated views of the same physical defect do not appear in both training and validation data.

2.5 Related work on textile and carbon-fibre defect inspection

Compared with mainstream computer vision domains, the literature specifically addressing defect segmentation on textiles and carbon-fibre materials remains relatively limited. This is particularly true for works that formulate the problem as a dense segmentation task rather than as image classification, object detection, or anomaly detection. Nevertheless, the existing studies provide useful methodological references and help position the present thesis within the current state of research.

A first line of related work is represented by studies on *textile defect segmentation*. Among these, the work of Xu et al. [4] is particularly relevant, as it addresses fabric defect segmentation using an improved Swin-Unet architecture and explicitly tackles some of the challenges associated with textured backgrounds and limited data. Their method combines a Transformer-based segmentation network with a fixed Gabor-filter preprocessing stage and introduces a loss function based on a weighted combination of focal loss and Dice loss. For these reasons, this paper was relevant not only as a related-work reference, but also as a practical source of ideas regarding loss design. Xu et al. explicitly report the use of the Cropped AITEX dataset and describe their weighted focal–Dice loss for handling data imbalance and difficult samples.

A second and even more directly related contribution is the recent work of Kandiyil et al. [5], which addresses defect localisation on carbon-fibre surfaces using deep-learning-based segmentation methods. This study is particularly relevant to the present thesis because it investigates semantic segmentation in the same industrial application domain and includes SegFormer among the main architectures under evaluation. In addition to SegFormer-based semantic segmentation, the authors also investigate instance segmentation using YOLOv11-seg and compare different sensing configurations based on polarization imaging under multiple illumination conditions. Their analysis reports that the additional polarization information provided only limited benefits compared with the standard monochrome sensor output.

The work is especially relevant because it shares with the present thesis several important characteristics, including the focus on carbon-fibre defect inspection, the use of Transformer-based segmentation architectures, and the investigation of different image representations for industrial sensing.

Despite these similarities, the two works differ in several important aspects. First, the considered defect taxonomy is substantially different. Kandiyil et al. group the anomalies into two broad categories (“carbon fibre defects” and “foreign objects”), whereas the present thesis considers a wider set of defect classes characterized by different spatial morphology and structural properties.

Second, the acquisition setup and sensing modalities are different. Kandiyil et al. rely primarily on grayscale imagery derived from a polarization camera, whereas the present work employs a proprietary industrial sensing system based on photometric stereo. The sensor acquires multiple observations under different illumination directions and produces several derived channels encoding physical surface properties such as fibre orientation and reflectance behaviour, rather than conventional grayscale or RGB information.

As a consequence, although both works are related at the level of task formulation and model family, they operate under substantially different sensing and data-representation conditions. Nevertheless, the study of Kandiyil et al. provides one of the closest currently available

references to the problem investigated in this thesis. The relevance of this work for the present thesis extends beyond the related-work perspective. Since the authors kindly shared the dataset used in their study, an additional exploratory cross-dataset transfer-learning experiment was later conducted using their industrial carbon-fibre dataset as an intermediate fine-tuning stage. This experiment will be discussed in Chapter 6. More generally, the work of Kandiyil et al. provides one of the closest currently available references to the problem considered in this thesis, both in terms of application domain and in terms of the investigated segmentation architectures.

In addition to the external literature discussed above, the present thesis also builds upon an earlier industrial study [6] developed in the same application context, together with the associated industrial dataset [7]. The previous work primarily addressed defect classification rather than semantic segmentation. Nevertheless, pixel-level annotations were already available because segmentation masks had been introduced as an auxiliary component to support the classification pipeline and to localize defective regions within the inspected material. However, since the original annotations were not created within a dedicated semantic segmentation framework, a systematic inspection of the dataset revealed several limitations with respect to the objectives of the present work. In particular, significant annotation inconsistencies were observed across different acquisitions of the same defect instance, multiple samples contained incomplete annotations, and the overall dataset size and variability were insufficient for robust supervised segmentation training. These observations motivated one of the central contributions of this thesis: the expansion and partial reconstruction of the dataset through a new controlled acquisition and annotation campaign specifically designed for leakage-safe semantic segmentation experiments, as described in Chapter 3.

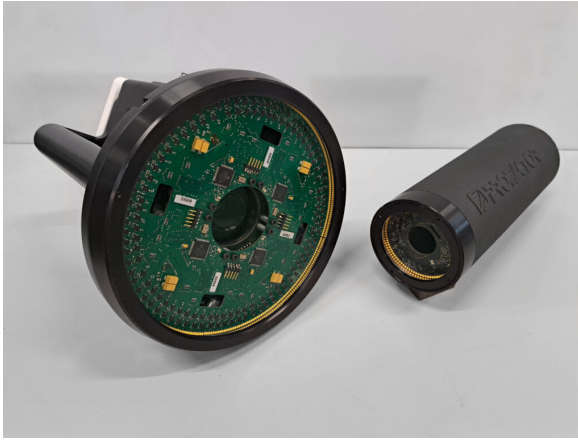
Overall, the existing literature confirms both the relevance and the difficulty of defect inspection on textured materials and carbon-fibre surfaces. At the same time, it also highlights that relatively few works address this problem through dense segmentation, and fewer still in conditions closely related to those considered here. This limited but growing body of work provides a useful reference framework for the present thesis, while also underscoring the need for careful attention to sensing modality, defect morphology, representation design, and evaluation methodology.

3 Dataset Analysis and Construction

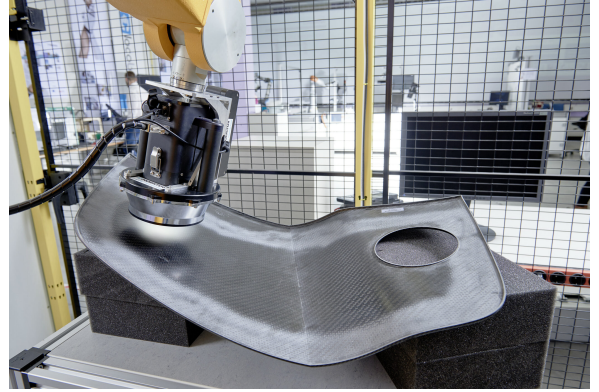
This chapter describes the data-related foundations of the work and explains how the final dataset used in this thesis was constructed. The discussion is organized around two complementary aspects. First, the industrial acquisition setting is introduced, including the sensing system used to collect the data and the sensor-derived representations available for inspection. Second, the original industrial dataset available at the beginning of the project is analysed in order to identify the limitations that made it unsuitable as the sole basis for a robust supervised segmentation study. Finally, the design and acquisition of a new dataset are presented, together with the annotation protocol, dataset statistics, grouped multi-view structure, and qualitative annotation rationale. The central argument of this chapter is that the main difficulty of the problem does not arise only from model selection, but also from the quality, consistency, and structure of the data itself. For this reason, dataset analysis and construction are not merely preparatory steps, but a core component of the thesis.

3.1 Industrial sensing system

The data considered in this work was acquired using the proprietary *Fscan* inspection system developed at PROFACTOR and employed within the framework of automated inspection of carbon-fibre textile materials in the European project *MC4 – Multi-level Circular Process Chain for Carbon and Glass Fibre Composites*. The *Fscan* sensor, shown in Figure 3.1, is a dedicated sensing platform designed for fibre-material inspection. Unlike standard RGB imaging systems, the *Fscan* system acquires information about the inspected surface through directional illumination. The sensor consists of a camera surrounded by multiple controllable illumination sources. During acquisition, different groups of LEDs are activated sequentially, producing several grayscale observations of the same material region under different lighting directions. This acquisition principle is closely related to photometric stereo. In classical photometric stereo, multiple images of the same surface are acquired from a fixed camera viewpoint while varying the direction of illumination. The changes in observed intensity across the image sequence provide information about the local interaction between light and surface geometry.



(a) Front view of the Fscan inspection sensor.



(b) Fscan system integrated in an industrial work-cell.

Figure 3.1: PROFACTOR Fscan inspection system used for data acquisition in this work.

In particular, the same surface point may appear brighter or darker depending on the orientation of the surface with respect to the active light source.

This principle is especially relevant for carbon-fibre materials. Carbon-fibre textiles exhibit strongly anisotropic reflectance behaviour: the amount of light reflected toward the camera depends heavily on the local orientation of the fibre bundles and on the direction of illumination. As a result, a single conventional image may not be sufficient to reveal all relevant structural information. By observing the same material region under multiple illumination directions, it becomes possible to emphasize orientation-dependent patterns, local changes in reflectance, and structural irregularities that may be associated with defects. In the present work, photometric stereo is not employed for explicit geometric reconstruction. Instead, the underlying acquisition principle is integrated into the Fscan sensing pipeline in order to generate multiple sensor-derived image representations emphasizing different reflectance and orientation properties of the material surface. During acquisition, the material is observed under multiple directional illumination conditions. The resulting observations are processed by the inspection software to produce feature channels that are more directly useful for fibre inspection than standard RGB or grayscale imagery. These representations encode complementary physical information related to fibre orientation, reflectance behaviour, and local surface structure.

Each acquisition is stored in the proprietary `.fscan` format, which contains both the raw directional illumination stack and the corresponding derived feature channels.

3.2 Sensor-derived representations

From the directional observations acquired by the *Fscan* system, several sensor-derived representations are computed in order to emphasize different physical properties of the carbon-fibre surface, including fibre orientation and reflectance behaviour.

Figure 3.2 shows an example of the raw directional observations acquired for a single material region.

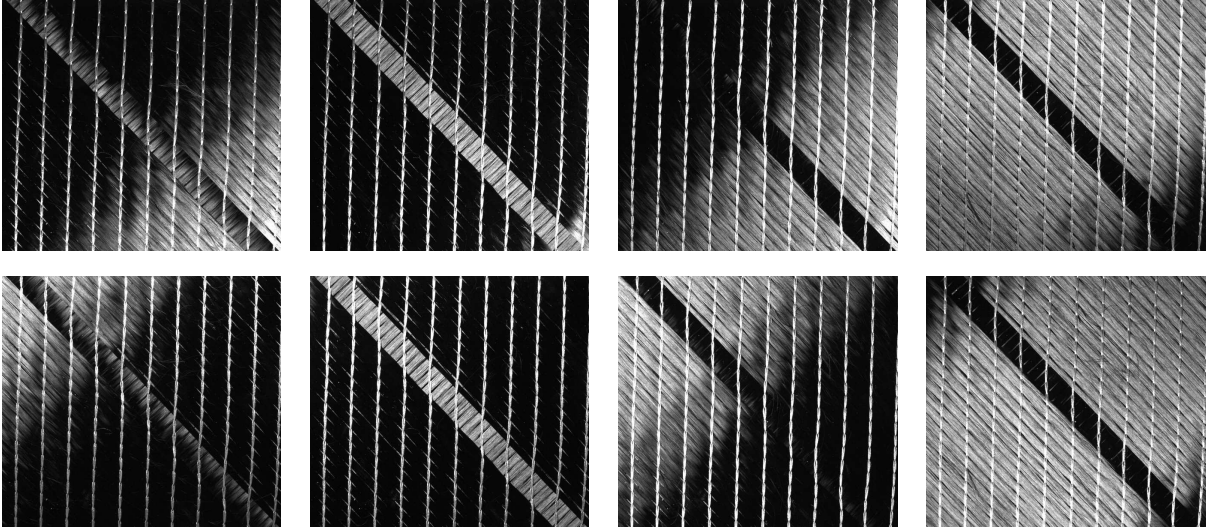


Figure 3.2: Example of the raw directional observations acquired for the same material region under different illumination directions. Variations in brightness and shadow patterns emphasize different surface and fibre characteristics depending on the illumination geometry.

The raw directional observations are subsequently processed by the inspection software to generate a set of sensor-derived feature representations that emphasize complementary physical properties of the carbon-fibre surface.

The channels considered in this thesis are:

- **azimuthal**, encoding the in-plane fibre orientation;
- **polar**, encoding the out-of-plane fibre orientation;
- **diffuse**, representing diffuse reflectivity;
- **specular**, representing specular reflectivity;
- **fibular**, providing a structurally intuitive visualization of the textile layout derived from the fibre reflection analysis.

The azimuthal channel intrinsically corresponds to a single-channel orientation representation whose values lie in the angular range $[-\pi, \pi]$. In the visualization shown in Figure 3.3,

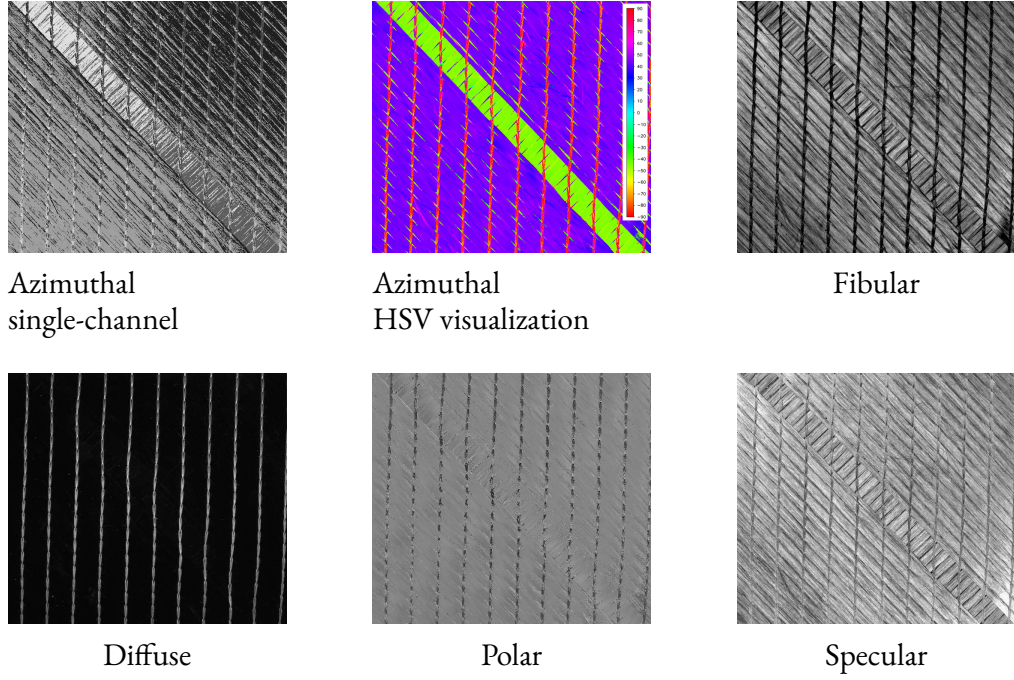


Figure 3.3: Example of the sensor-derived representations generated by the Fscan processing pipeline from the raw directional illumination observations. The azimuthal channel encodes local fibre orientation through angular values and is shown both as a normalized single-channel representation and as an HSV-based visualization used to highlight orientation differences. The remaining channels emphasize complementary physical properties of the material surface, including fibre orientation, fibre reflectivity, and reflectance behaviour under directional illumination.

the angular values are mapped to HSV colors in order to make local orientation variations easier to interpret visually. This HSV conversion is introduced only for visualization purposes and is not used directly as input to the neural network. During model training, the azimuthal information is instead represented as a single normalized channel. Different normalization strategies were investigated, including direct normalization to $[0,1]$ and modulo-based mappings designed to preserve orientation equivalence. These representation choices are discussed in detail later in Chapter 5.

Different defect types may become more evident in different channels. Some anomalies are primarily associated with reflectance variations, whereas others are more clearly characterized by disruptions of fibre orientation. Consequently, the selection and representation of the sensor-derived channels become an integral part of the learning pipeline rather than a fixed pre-processing choice. The normalization and combination strategies adopted for these channels are discussed later in Chapter 5. In the present chapter, the channels are introduced primarily to describe the structure of the acquired data and the annotation workflow used during dataset construction.

3.3 Limitations of the original dataset

As discussed in Chapter 2, the original industrial dataset was not specifically designed for supervised semantic segmentation. For this reason, a detailed analysis of its annotation structure and acquisition characteristics was conducted in order to identify the main limitations motivating the construction of a new dataset for the present work.

A first aspect concerned the annotation structure. Defects belonging to the same semantic class were not always delineated with the same spatial extent or level of precision. In some cases, the annotation followed the visible defect boundary closely, whereas in others it also included a larger surrounding region of material. Since the dataset had not been constructed specifically for a dedicated segmentation study, this degree of variability is understandable, but it becomes relevant when the goal is to provide consistent pixel-level supervision to a segmentation model. A related issue emerged when comparing repeated acquisitions of the same physical defect. In several cases, the same underlying defect was imaged multiple times under different positions or orientations with respect to the sensor. This grouped acquisition structure is not a drawback in itself; on the contrary, it is a useful way to increase the number of available observations and to capture appearance variability induced by pose and sensing conditions. However, it also makes annotation variability easier to observe, because the same physical defect can be compared across multiple related acquisitions. In some cases, these repeated views were annotated with slightly different spatial extent, partial delineations, or non-uniform geometric coverage.

A further aspect concerned the completeness of the available annotations. Some images contained more than one visible defect, while only a subset of those defective regions was explicitly annotated. This behaviour is consistent with the original formulation of the dataset as a single-label multi-class classification problem, where the primary objective was to assign a dominant defect category to the image rather than to exhaustively delineate all visible defect instances. In a supervised segmentation setting, this becomes important because any unlabeled defective region is implicitly interpreted as background during training. As a result, the issue is not simply one of annotation noise, but of partial supervision that may affect both optimization and evaluation.

Taken together, these observations suggested that the original dataset should be interpreted with care in the context of the present thesis. The main issue was not that the dataset was unusable in absolute terms, but that it had not been designed specifically for the requirements of a systematic supervised semantic segmentation study.

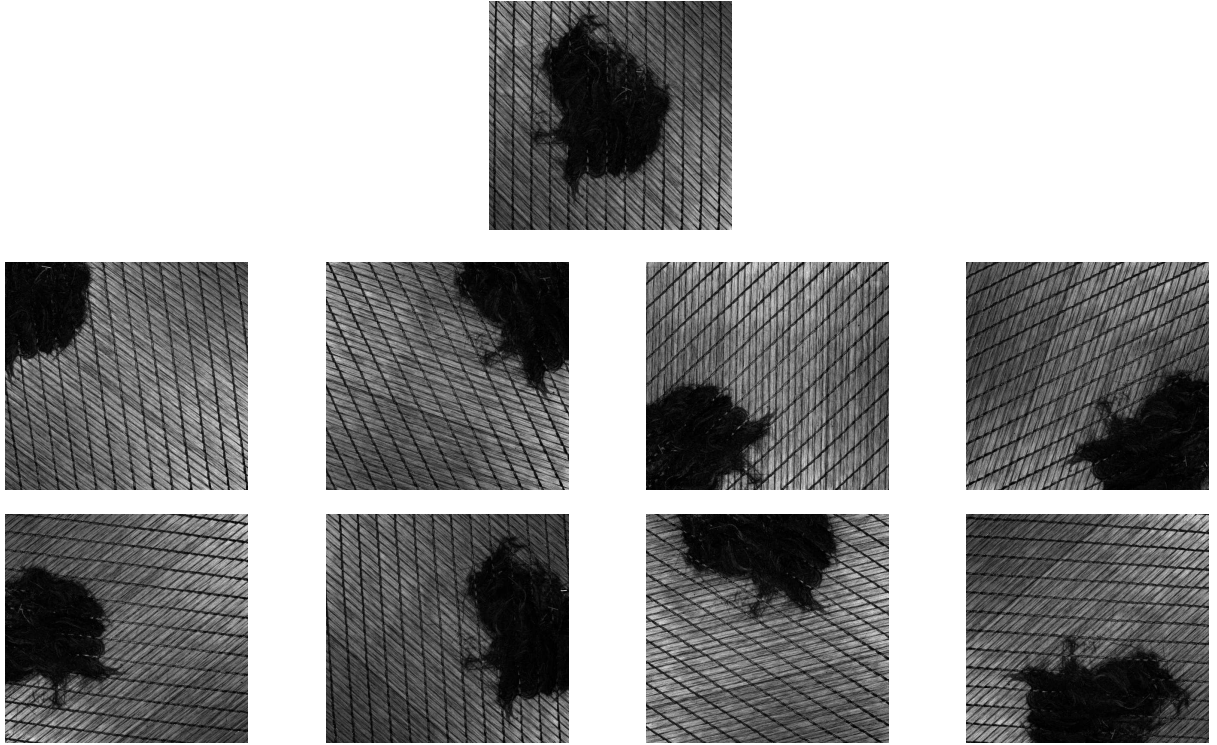


Figure 3.4: Acquisition convention adopted for the new dataset. The same physical defect is acquired multiple times under controlled variations of position and orientation with respect to the sensor reference frame.

3.4 Dataset reconstruction

Motivated by the limitations of the original data, a new data acquisition campaign was initiated with the explicit goal of producing a dataset suitable for supervised segmentation. The design of this dataset addresses the limitations identified in the original industrial data by enforcing:

- consistent annotation guidelines across all defect classes;
- controlled acquisition protocols to reduce variability unrelated to defects;
- explicit grouping of correlated acquisitions in order to support leakage-safe evaluation protocols.

During dataset acquisition, each physical defect was observed multiple times under controlled variations of position and orientation with respect to the sensor reference frame. This strategy was adopted in order to increase the diversity of the acquired samples while preserving the identity of the underlying physical defect instance.

Figure 3.4 illustrates the acquisition convention adopted throughout the data collection process.

The resulting dataset therefore contains grouped acquisitions corresponding to multiple views of the same physical defect. This grouped structure was explicitly preserved in the dataset organization and later taken into account during the definition of the evaluation methodology discussed in Chapter 4.

3.4.1 Authenticity of defect samples

An important design choice during dataset construction was to acquire only *naturally occurring defects* present in the inspected carbon-fibre preforms. No attempt was made to artificially introduce defects such as manually cutting fibres, inserting contaminants, or intentionally altering the textile structure. Several academic datasets adopt controlled procedures to synthetically create defects in order to balance class frequencies or increase the number of available samples. In the present work, this approach was intentionally avoided. In addition to preserving realistic defect characteristics, the available industrial material samples were also part of parallel activities and therefore could not be modified exclusively for dataset generation purposes. While synthetic defect creation can simplify dataset construction and facilitate more balanced training scenarios, it may also produce defect appearances that differ from those encountered in real manufacturing conditions. In contrast, the dataset presented in this work was collected exclusively from authentic material samples obtained during inspection. As a result, the distribution of defect types reflects the availability and variability of real industrial samples rather than a deliberately balanced acquisition scheme. Since the collected preforms originated from different industrial sources and production contexts, the dataset should not be interpreted as representing a single manufacturing process. Some defect categories appear frequently, whereas others occur only rarely.

Although this choice introduces a degree of class imbalance at the annotation level, it provides a more faithful representation of the inspection conditions encountered in industrial practice.

Furthermore, the primary learning task in this work is formulated as *binary defect segmentation* (defect versus background). Consequently, the presence of long-tailed distributions among individual defect categories has a limited impact on the training objective, since all defect types are ultimately collapsed into a single foreground class.

3.4.2 Annotation methodology

A central design choice of the new dataset is the adoption of *instance-level* segmentation annotations. All samples were annotated using CVAT (Computer Vision Annotation Tool) [8], employing polygon-based annotations to precisely delimit defect regions. Each polygon corre-

sponds to a single defect instance and is associated with a specific defect class. This annotation strategy provides maximal flexibility: the dataset can be directly used for instance segmentation and, through a simple binarization of the instance masks, also for semantic segmentation.

In this work, all defect instances are collapsed into a single foreground class (*defect*) for defect/background semantic segmentation. However, the original multi-class, instance-level annotations are preserved, enabling future experiments on multi-class semantic segmentation or instance segmentation without re-annotation.

The polygon tool allows the annotator to manually trace the exact spatial extent of a defect, which is particularly important for carbon-fibre materials, where defect boundaries are often irregular, elongated, or topology-dependent and do not align with simple geometric primitives. Compared to bounding-box annotations, polygon-based segmentation provides a more faithful representation of defect morphology and supports pixel-accurate supervision, which is critical for learning fine-grained spatial cues. To further improve the reliability and consistency of the annotations, the official **PROFACTOR Fscan software** was used alongside CVAT during the annotation process. In particular, the Fscan software was employed to inspect the derived feature representations produced by the sensor, including the five channels described in the previous chapter. Visual inspection of these channels (azimuthal, fibular, specular, diffuse, and polar) provided complementary information that is not always apparent in single-channel visualizations. This multi-channel inspection was especially useful for identifying subtle defects and for accurately delineating defect boundaries in cases where the contrast in one channel was weak but clearly visible in another. The availability of sensor-native visualization tools facilitated more consistent distinction between true defects and normal texture variations, improving the reliability of polygon boundary delineation across samples and defect classes.

Annotations generated in CVAT can be exported in multiple standard formats, including YOLO segmentation, COCO, and dense pixel-wise masks. This interoperability simplifies integration with different training pipelines, supports reproducibility, and facilitates dataset reuse beyond the scope of this work.

3.5 Practical channel preference during annotation

In the majority of cases, defect boundaries were delineated using the *fibular* channel. This representation closely resembles a uniformly illuminated grayscale view of the textile surface and provides intuitive visual correspondence to the physical material structure. For most defect types, it offered stable contrast and clear boundary perception. However, for certain defect categories, alternative channels were preferentially used to improve boundary visibility. In particular:

- the *diffuse* channel was advantageous when local reflectivity variations were more informative than structural texture;
- the *azimuthal* channel was used when defects manifested primarily as orientation discontinuities in the fibre layout.

From an annotation perspective, the practical usefulness of each channel depends on how clearly defect boundaries emerge in the corresponding representation. Since different channels encode distinct physical surface properties, including fibre orientation and reflectivity behaviour, their usefulness for annotation varies depending on the defect morphology and the material’s optical behaviour.

3.6 Dataset statistics and defect taxonomy

The final dataset contains **417 samples** acquired across **58 distinct physical defect instances**. Each sample corresponds to a single acquisition of a material region containing one or more defects. Owing to the grouped multi-view acquisition protocol described in the previous sections, multiple samples may therefore correspond to different observations of the same physical defect instance. No fully defect-free samples are included in the annotated defect dataset reported in the following statistics. The dataset statistics therefore refer only to acquisitions containing one or more annotated defect regions. However, a separate auxiliary pool of non-defective material scans was available and was used only for data-augmentation purposes, as described in Section 5.6. These auxiliary background scans are not included in the defect dataset statistics.

Although the annotations are organized into multiple defect categories, the segmentation task considered in this thesis is formulated as a **binary segmentation problem** (*defect vs. background*), where all defect classes are collapsed into a single foreground label. Nevertheless, preserving the original defect taxonomy remains important for dataset analysis and for understanding the variability of the considered anomalies. At pixel level, the dataset exhibits a strong foreground-background imbalance, with approximately **4.97%** of the pixels across the entire dataset belonging to defective regions.

The dataset includes **nine defect classes**, covering a representative subset of anomalies commonly encountered in textile-based carbon-fibre preforms. These defects exhibit substantially different geometric structure, spatial extent, and visual appearance, thereby increasing the complexity of the segmentation task. A brief description of each defect category is provided in Table 3.1.

Overall, the dataset contains **850 annotated defect instances**. The per-class instance distribution is reported in Table 3.2.

Class	Description
Missing stitch	Absence of a binding stitch, resulting in locally unsecured fibre tows.
Gap	Unintended spacing between adjacent fibre tows, exposing the substrate.
Fuzzball	Localized accumulation of loose fibres forming a small fibre agglomerate on the surface.
Drilled tow	Fibre tow damaged or perforated, typically caused by mechanical intrusion.
Weaving error	Structural deviation from the intended weave pattern, such as misplaced or crossed tows.
Fuzz yarn	Extended loose fibres protruding from a yarn, often elongated rather than compact.
Distortion	Global or local misalignment of fibre tows, resulting in curvature or waviness of the pattern.
Foreign object	Presence of non-carbon material embedded in or resting on the textile surface.
Broken tow	Partial or complete rupture of a fibre tow, interrupting load-carrying continuity.

Table 3.1: Defect classes present in the dataset and their description.

Class	Instances	Images containing class
Missing stitch	79	66
Gap	188	113
Fuzzball	93	65
Drilled tow	54	54
Weaving error	56	56
Fuzz yarn	271	57
Distortion	84	54
Foreign object	9	9
Broken tow	16	16

Table 3.2: Number of annotated instances and images per defect class.

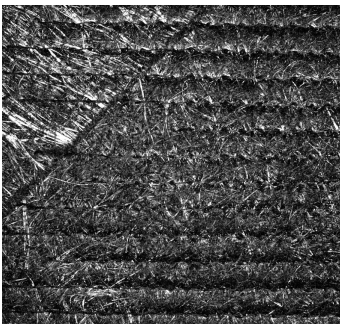
It can be observed that *fuzz yarn* is the most frequent class in terms of instance count (271 instances), whereas *foreign object* is the rarest category (9 instances).

Beyond class imbalance, the dataset also exhibits substantial variability in defect morphology and spatial extent. Some defects appear as small localized anomalies (e.g., *drilled tow* or *broken tow*), whereas others may occupy large continuous regions of the image (e.g., large-scale *distortion* patterns). In addition, certain categories present compact structures, while others exhibit elongated or highly irregular geometries. This variability increases the difficulty of the segmentation problem, since the model must simultaneously detect fine local defects and broader structural anomalies across multiple spatial scales. Furthermore, the number of defect instances per image varies considerably, and the per-image defect coverage ranges from 0.06%

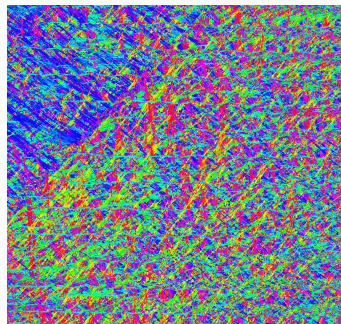
up to 34.83% of the image area. This wide variability in defect density and spatial occupancy further contributes to the heterogeneity of the dataset and motivates the need for robust multi-scale segmentation approaches capable of handling highly diverse defect configurations.

3.7 Recycled material subset

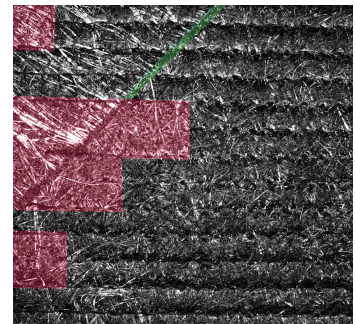
In addition to variability in defect morphology and spatial extent, the dataset also contains a small subset of samples acquired from recycled carbon-fibre material. These samples differ substantially from the standard woven textile specimens that constitute the majority of the dataset, introducing an additional source of visual and structural variability. Compared to standard textile preforms, the recycled material exhibits a significantly less regular fibre structure, characterized by fragmented fibre bundles, irregular spacing between tows, and locally varying fibre orientations. In the fibular representation, this produces a highly textured background containing numerous bright fibre fragments and small reflectance variations, often reducing the visual contrast between defective and non-defective regions. Similarly, the azimuthal orientation channel exhibits strong local variability, since the fibre orientations no longer follow a regular periodic pattern. These characteristics make the delineation of defect boundaries considerably more difficult, even during manual annotation. Figure 3.5 illustrates a representative example of this material type together with the corresponding annotation. Only four defect instances belonging to this recycled-material subset were available in the collected dataset. Consequently, the model is exposed to very limited examples of this material domain during training. When such samples appear in the validation set, they often produce substantially lower segmentation performance compared with standard woven textile samples.



Sample 1 – Fibular



Azimuthal (HSV)



Annotation overlay

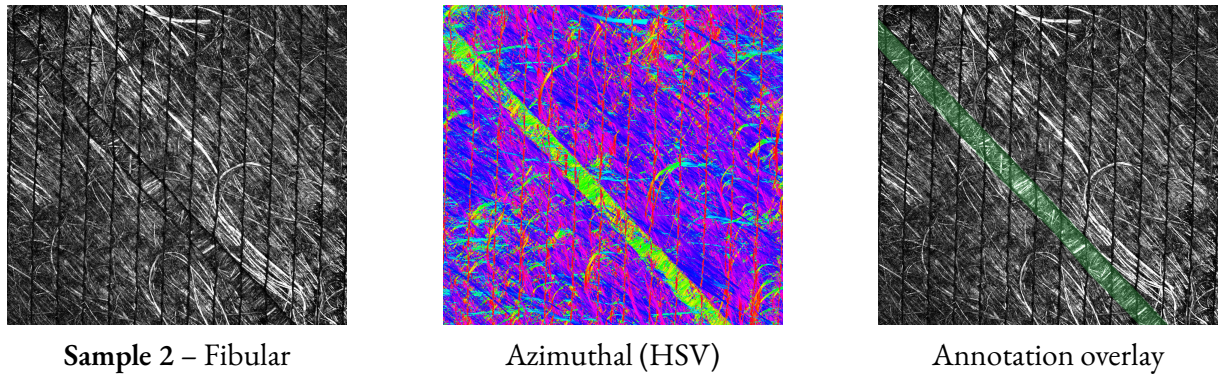


Figure 3.5: Compared with standard woven textile samples, the recycled material exhibits highly irregular fibre arrangements and reduced structural regularity, making defect delineation considerably more difficult. The first sample contains both a missing stitch (red) and a gap (green), illustrating the complex defect configurations that may appear in recycled material.

3.8 Qualitative inspection and annotation rationale

In addition to the quantitative statistics presented previously, it is important to analyse the visual appearance and spatial structure of the defects present in the dataset. The following sections therefore present representative examples for each defect category together with the annotation criteria adopted during polygon delineation.

For each defect class, the discussion highlights the visual characteristics used to identify the defect, the rationale behind the annotated spatial extent, and the conventions followed to maintain consistency across samples. This qualitative inspection provides additional insight into the morphological variability of the dataset and clarifies the practical decisions involved in the annotation process.

3.8.1 Missing stitch

Missing stitch defects correspond to the absence of a binding stitch, resulting in locally unsecured fibre tows. Figure 3.6 shows representative examples together with the visualizations used during annotation and the resulting overlays.

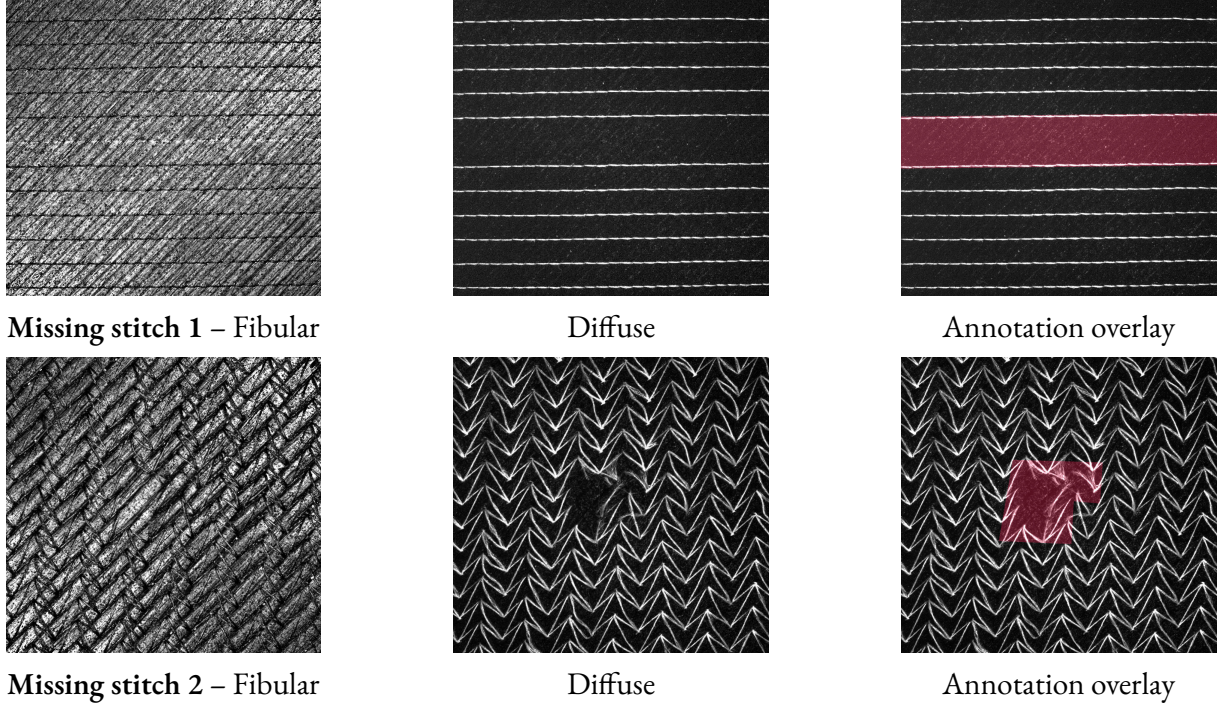


Figure 3.6: Composite visualization of missing stitch defects. Each row corresponds to a defect instance, while columns show the fibular representation, the diffuse channel, and the final annotation overlay.

Channel used for annotation. For this defect type, the diffuse channel provided clearer contrast between secured and unsecured fibre regions. Although the fibular representation offers a structurally intuitive view, the reflectance variations emphasized by the diffuse channel facilitated more accurate boundary delineation for this material instance. For this reason, the diffuse representation was adopted as the primary reference during polygon delineation for this defect category.

Annotation rationale and protocol refinement. An important observation emerging during dataset construction was that the segmentation performance for missing stitch defects was highly sensitive to the adopted annotation protocol. In an initial annotation attempt, polygons were restricted to the thin missing-stitch structure itself, producing extremely small foreground regions (Fig. 3.7(a)).

Under this representation, the supervision signal associated with the defect occupied only a negligible fraction of the image area. A plausible consequence is that the optimization problem becomes highly unbalanced, since predicting background over the entire image may produce a lower overall loss than attempting to localize a very thin structure with uncertain boundaries. To mitigate this issue, the annotation protocol was revised by expanding the annotated region to include the area between the two adjacent stitches surrounding the missing structure (Fig. 3.7(b)). This modification remains semantically consistent with the physical interpretation of the defect. In practice, a missing stitch does not only correspond to the thin absent binding structure itself, but also to the locally unsecured region between the adjacent stitched areas. Consequently, the expanded annotation provides a representation that is more aligned with the operational manifestation of the defect while simultaneously producing a stronger and more stable supervision signal during training. This choice is also aligned with practical industrial inspection requirements, where missing a defect is generally considered more critical than slightly overestimating its spatial extent.

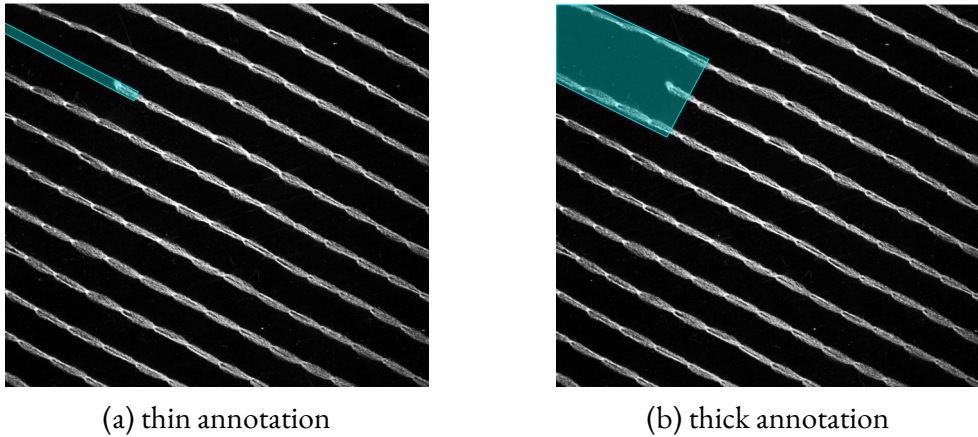


Figure 3.7: Annotation protocol difference.

It is important to note that this form of mask expansion was adopted specifically because it remained semantically consistent with the physical interpretation of the defect. The expanded region still corresponds to the portion of material affected by the absence of the binding structure. In contrast, applying arbitrary enlargement to other defect classes would incorrectly include undamaged material and would alter the semantic meaning of the annotation. Consequently, this strategy was used selectively and only when the expanded region could still be considered physically and operationally representative of the defect itself.

3.8.2 Gap

Gap defects correspond to a separation between adjacent fibre tows, which exposes the underlying substrate. This separation typically introduces two observable effects: the presence of the exposed gap itself and a deformation of the surrounding fibre layout caused by the displacement of neighbouring tows.

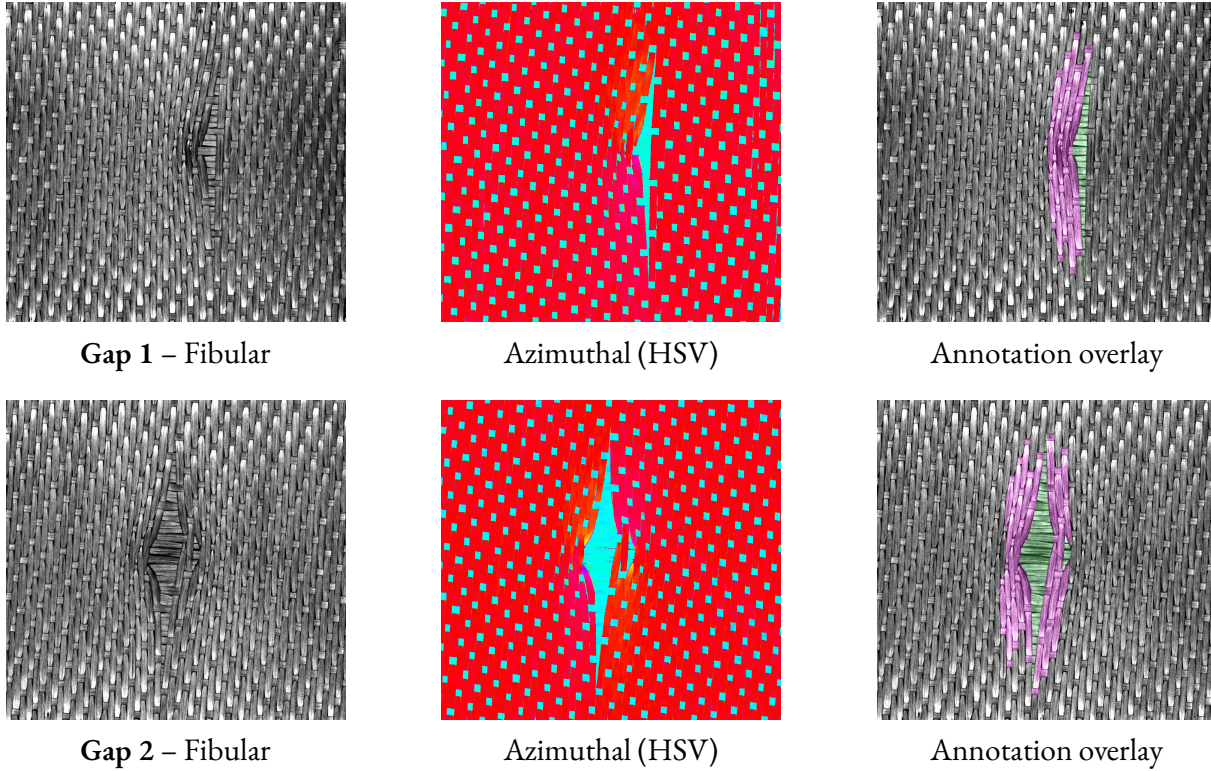


Figure 3.8: Composite visualization of gap defects. Each row corresponds to a defect instance, while columns show the fibular representation, the azimuthal orientation visualization rendered in HSV, and the final annotation overlay. The annotation distinguishes the gap region from the surrounding fibre distortion induced by the separation between adjacent tows.

Channel used for annotation. Initial annotation was primarily performed using the fibular representation, which provides a grayscale visualization of the fibre reflectivity response and textile structure. In this channel, the exposed substrate associated with the gap appears with a different reflectivity response that differs from the surrounding fibre tows.

However, the azimuthal orientation channel often provided a clearer indication of the true defect boundary, since it encodes local fibre orientation and therefore produces a sharp discontinuity between intact fibre regions and the exposed substrate. In the HSV visualization, this transition appears as an abrupt color variation that facilitates more precise boundary delineation. In addition to the exposed gap itself, the separation between adjacent tows frequently

produces local deformation of the surrounding fibre layout. These distortions manifest as irregularities in fibre orientation and spacing around the defect region. For this reason, the distorted fibre area surrounding the gap was annotated separately from the exposed substrate region itself.

3.8.3 Fuzzball

Fuzzball defects correspond to localized accumulations of loose fibres on the textile surface. Their appearance is primarily driven by changes in surface texture and reflectance rather than by fibre orientation discontinuities. Figure 3.9 presents two representative examples together with the corresponding annotation overlays.

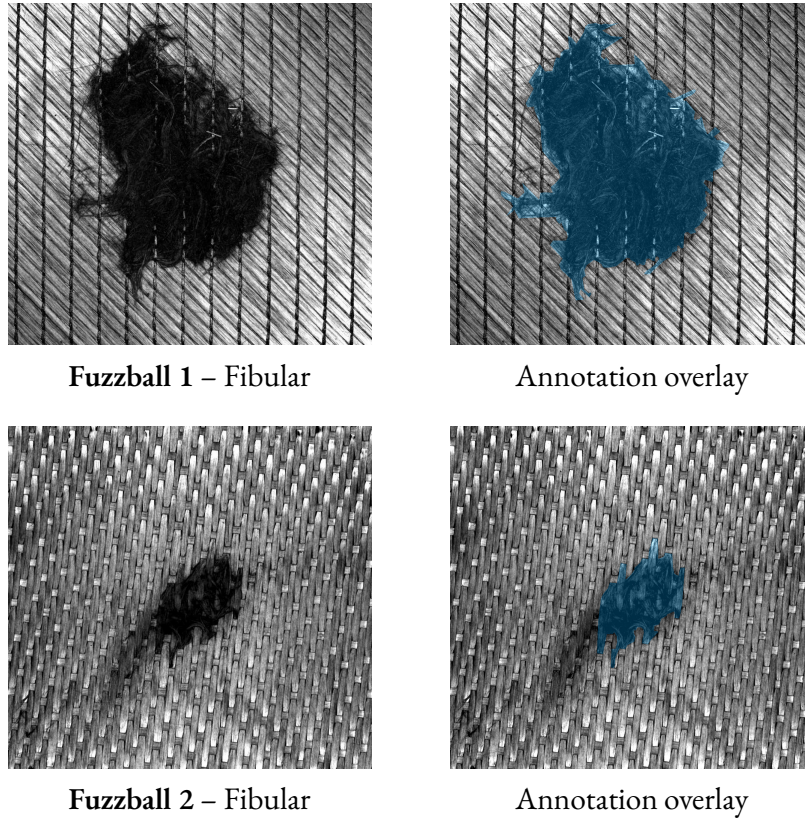


Figure 3.9: Composite visualization of fuzzball defects. Each row corresponds to a defect instance. The fibular representation (left) was used during annotation, while the images on the right show the final polygon overlays.

Channel used for annotation. Fuzzball defects were annotated directly on the fibular representation. Unlike defects such as gaps or weaving errors, fuzzballs do not produce strong discontinuities in fibre orientation. Instead, they appear as localized fibre agglomerations that disrupt the regular textile texture and modify the local reflectance pattern.

In the fibular channel, these defects typically appear as darker and more irregular regions with increased visual clutter compared with the surrounding textile structure, as shown in Fig. 3.9. Since the dominant visual cue is associated with texture and reflectance variations rather than orientation changes, orientation-based channels did not provide substantial additional boundary information for this defect category.

3.8.4 Drilled tow

Drilled tow defects correspond to localized damage or perforation of a fibre tow, typically caused by mechanical intrusion. Figure 3.10 shows two representative examples together with the final annotation overlays.

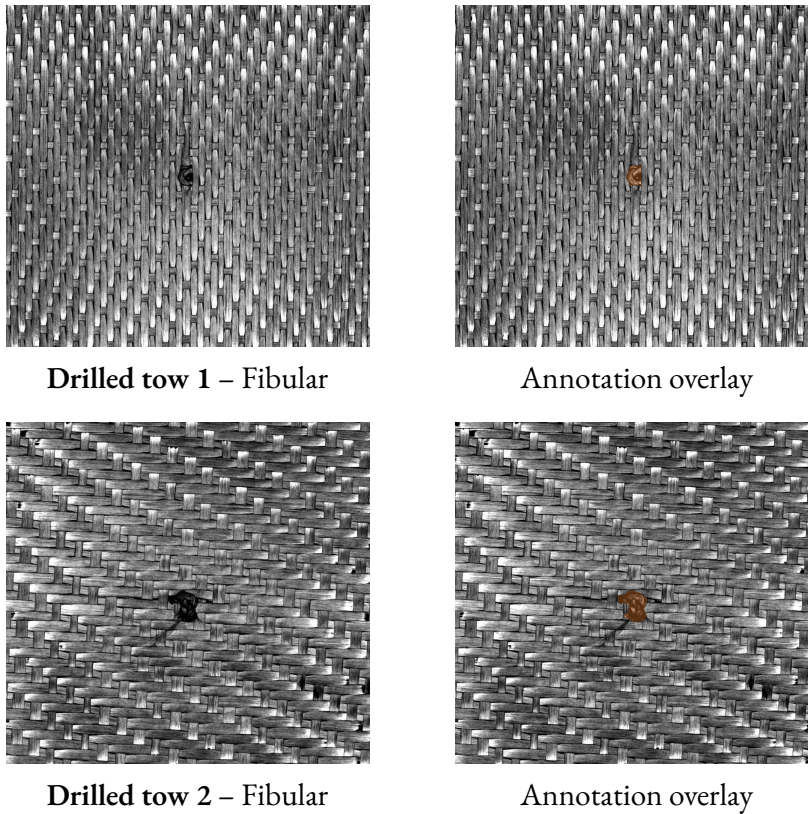


Figure 3.10: Composite visualization of drilled tow defects. Each row corresponds to a defect instance. The fibular representation (left) was used during annotation, while the images on the right show the final polygon overlays.

Channel used for annotation. Annotations for drilled tow defects were performed on the fibular representation. In this defect category, the primary cue is a localized disruption of the tow appearance (missing/damaged fibres, irregular texture, and local contrast changes) rather

than a characteristic orientation discontinuity. The fibular view therefore provides an intuitive and sufficiently sharp visualization for placing polygon vertices along the damaged tow boundary.

3.8.5 Weaving error

Weaving errors correspond to structural deviations from the intended textile pattern, such as locally misplaced or crossed tows. In this defect category, the visual evidence is often subtle in reflectance-based representations, and the defect extent is not always sharply delimited. Figure 3.11 shows two representative samples together with the azimuthal visualization and the resulting annotation overlays.

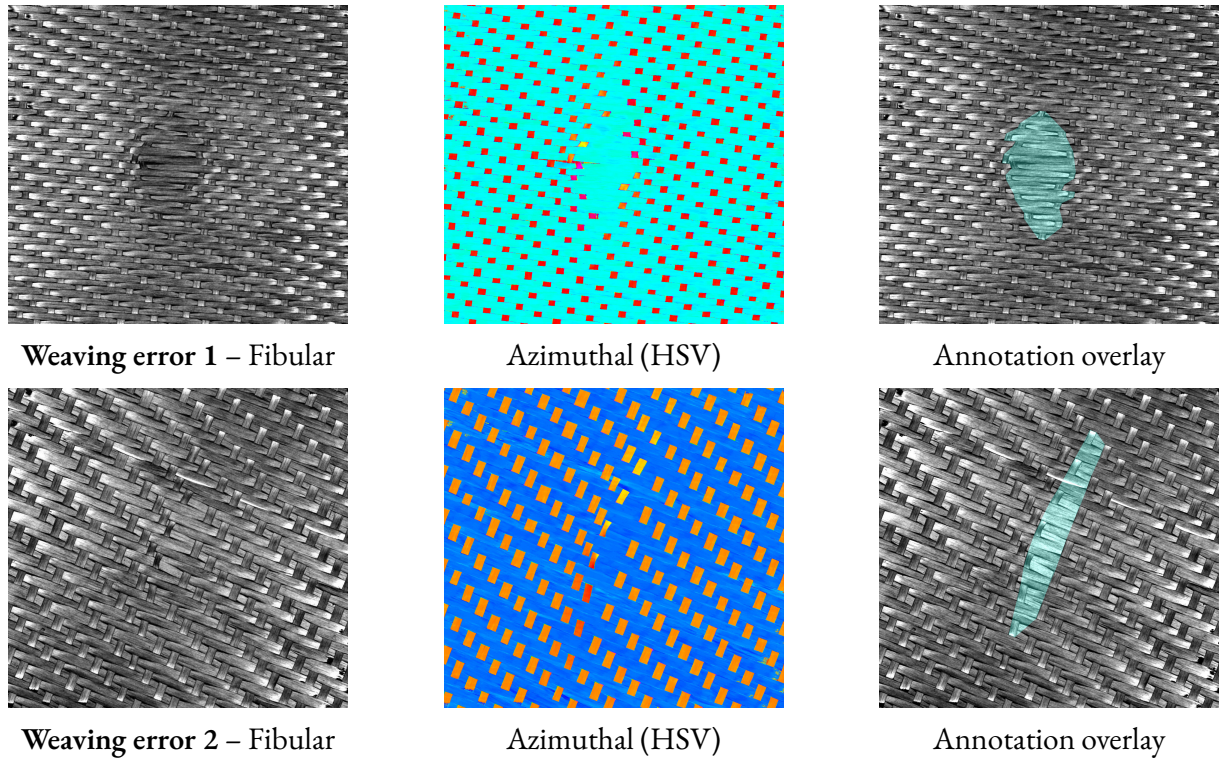


Figure 3.11: Composite visualization of weaving error defects. Each row corresponds to a defect instance; columns show the fibular representation, the azimuthal channel visualized in HSV, and the final annotation overlay. Weaving errors are primarily identified through local changes in tow spacing and orientation irregularities highlighted in the azimuthal visualization.

Channel used for annotation. Weaving errors were annotated through joint inspection of the fibular and azimuthal representations, with primary reliance on the latter. Since this representation directly encodes in-plane fibre orientation, misaligned or crossed tows appear as coherent regions whose hue differs from the surrounding regular pattern. In contrast, reflectivity-

based channels alone often do not provide sufficiently stable boundary cues for this defect category.

Annotation protocol and boundary definition. The most reliable indicator of a weaving error in the considered samples was an increase in spacing between neighbouring tow elements combined with localized orientation deviations visible in the HSV visualization. Because these transitions are often gradual rather than sharply localized, the defect boundary was defined conservatively by enclosing neighbouring regions whose hue deviated noticeably from the dominant background pattern. In practice, polygon vertices were placed along the envelope of the tow elements exhibiting coherent orientation deviations, providing a repeatable annotation criterion across samples.

Distinction from gap defects. Although both *gap* and *weaving error* defects may involve increased spacing between adjacent tows, their visual manifestation differs substantially. In gap defects, the separation typically exposes the underlying substrate, producing a clearly delimited region with strong contrast in reflectivity-based visualization. In weaving errors, the spacing irregularity does not necessarily expose the substrate, and the fibular appearance may therefore remain similar to the surrounding material. Consequently, orientation-based cues derived from the azimuthal channel were essential for consistent delineation of weaving errors.

3.8.6 Fuzz yarn

Fuzz yarn defects correspond to loose fibres protruding from an individual yarn, resulting in localized fraying of the tow structure. Unlike fuzball defects, which typically form compact fibre accumulations, fuzz yarn defects exhibit a more elongated structure aligned with the yarn direction and are generally confined to a single tow. Figure 3.12 shows two representative examples together with the corresponding annotation overlays.

Channel used for annotation. Fuzz yarn defects were annotated directly on the fibular representation. In this defect category, the dominant visual cue is a localized disruption of the yarn texture caused by loose fibres emerging from the tow structure. These fibres typically produce an elongated darker region aligned with the yarn direction, often with increased local irregularity compared with the surrounding textile pattern.

Since the defect primarily affects local texture and reflectance rather than fibre orientation, orientation-based channels such as the azimuthal representation did not provide substantial additional information for boundary delineation.

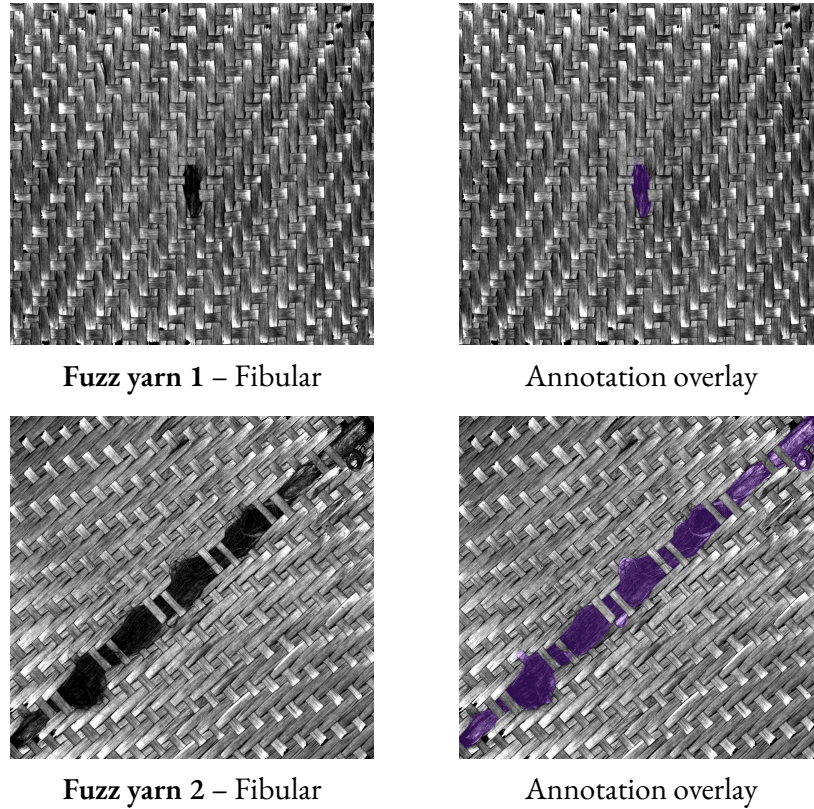


Figure 3.12: Composite visualization of fuzz yarn defects. Each row corresponds to a defect instance. The fibular representation (left) was used during annotation, while the images on the right show the resulting polygon overlays delineating the elongated regions affected by protruding fibres along the yarn direction.

3.8.7 Distortion

Distortion defects correspond to local deviations from the regular fibre layout of the textile. Instead of following the nominal weave trajectory of the material, the affected fibres appear bent, displaced, or locally misaligned with respect to the surrounding structure. Figure 3.13 shows two representative examples together with the resulting annotation overlays.

Channel used for annotation. Distortion defects were annotated primarily using the fibular representation. In this channel, the periodic fibre layout of the textile is clearly visible, making it possible to visually follow the nominal trajectory of the fibre tows. Regions where fibres deviate from this expected arrangement are therefore relatively easy to identify.

Annotation protocol. The annotation criterion adopted for this defect category was to include all fibre regions that visibly deviate from the regular orientation and spatial arrangement of the surrounding material. In practice, the annotated area corresponds to the envelope of the fibre segments that no longer follow the nominal textile structure.

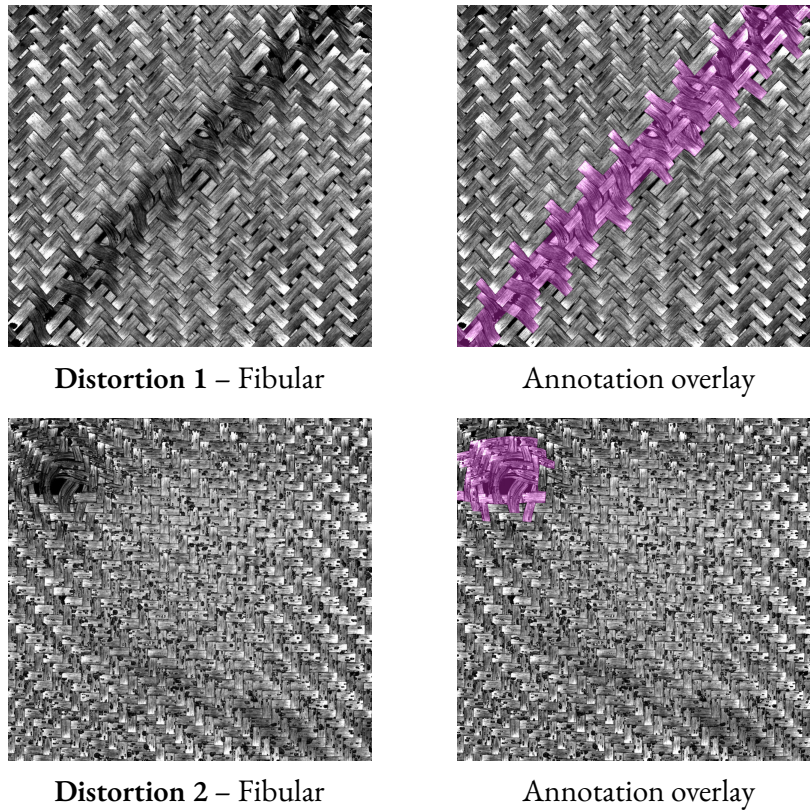


Figure 3.13: Composite visualization of distortion defects. Each row corresponds to a defect instance. The fibular representation (left) was used during annotation, while the images on the right show the final polygon overlays marking the region where fibres deviate from the regular textile structure.

Although distortion defects may appear together with other structural irregularities such as gaps, they are not necessarily dependent on them and may also occur independently.

3.8.8 Foreign object

Foreign object defects correspond to the presence of material that does not belong to the carbon-fibre textile structure, such as dust particles, contaminants, or small fragments introduced during handling or manufacturing.

Channel used for annotation. The annotation for this defect was performed using the fibular representation. In this channel, the foreign object appears as a localized region whose reflectance characteristics differ from the surrounding fibre structure, making the affected area visually distinguishable from the background material.

Annotation protocol and limitation. In the considered sample, the contaminant was attached to the material using a small piece of transparent tape. While the foreign particle itself

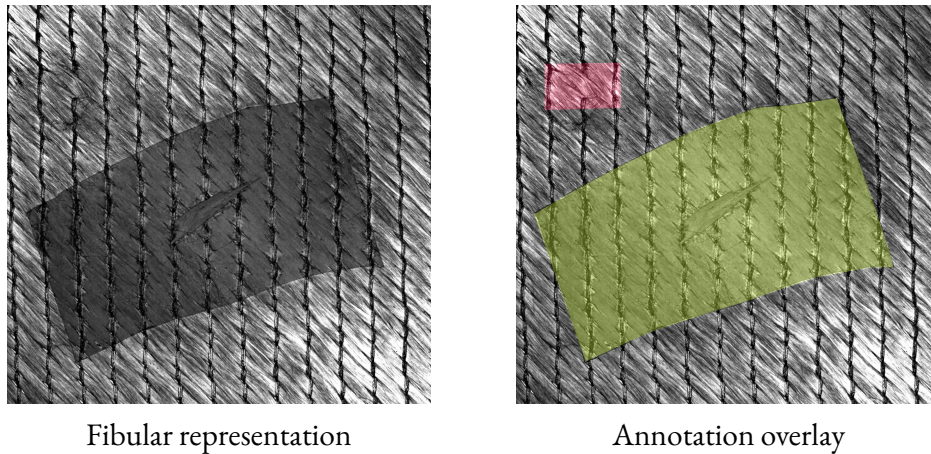


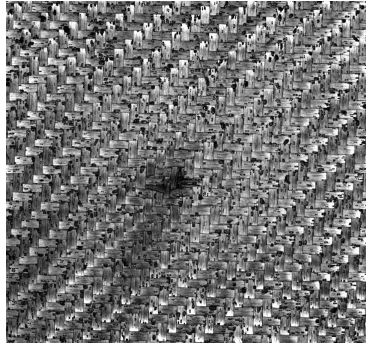
Figure 3.14: Example of a foreign object defect. The fibular channel reveals a localized region whose appearance differs from the surrounding fibre structure. The right image shows the resulting polygon annotation used during dataset construction. The sample also contains a small missing stitch, which is annotated as an additional defect region in the overlay.

is visible in the azimuthal representation, the contour of the tape appears much more clearly in the fibular channel. This created an ambiguity during annotation, since the visible boundary in the inspection data did not perfectly coincide with the physical extent of the actual contaminant. Following the annotation policy adopted throughout the dataset, polygon boundaries were placed according to the most stable and clearly identifiable contour observable in the channel used during annotation. Consequently, the annotation enclosed the entire tape region rather than only the contaminant itself. Although the transparent tape is not part of the actual foreign object and would not normally appear in a real inspection scenario, this choice provided a more consistent and reproducible annotation criterion for this particular sample. As a result, the annotated region is larger than the contaminant itself. For future acquisitions of this defect category, the use of adhesive tape should be avoided entirely in order to preserve the true visual appearance of the defect under realistic deployment conditions.

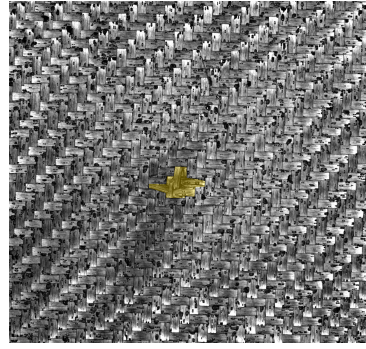
3.8.9 Broken tow

Broken tow defects correspond to the rupture of one or more fibre bundles within the textile structure. This discontinuity produces a localized disruption of the regular fibre layout and may also induce small deformations in the surrounding material.

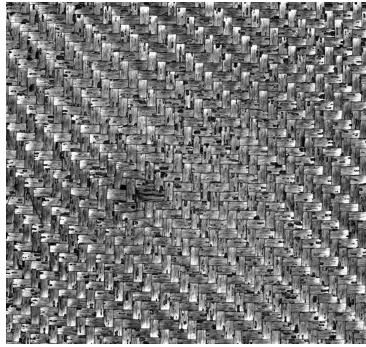
Channel used for annotation. Broken tow defects were annotated directly on the fibular representation. In this channel, the rupture of the fibre bundle appears as a localized interruption of the regular textile structure, making the damaged region visually distinguishable from the surrounding material.



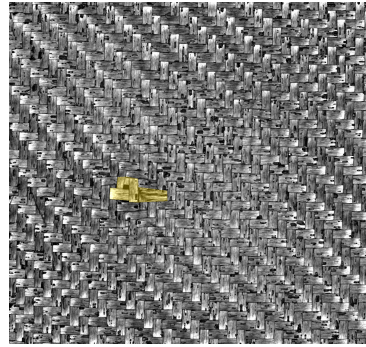
Broken tow 1 – Fibular representation



Annotation overlay



Broken tow 2 – Fibular representation



Annotation overlay

Figure 3.15: Examples of broken tow defects. Each row corresponds to a defect instance. The fibular representation provides a clear visualization of the disrupted fibre bundle, while the overlay images show the final polygon annotations used during dataset construction.

Annotation protocol. The annotated region includes the portion of the tow where the fibre bundle appears visibly broken or interrupted. Polygon boundaries were placed along the irregular contour of the damaged fibre structure in order to capture the spatial extent of the defect.

3.9 Dataset summary

The previous sections described the acquisition procedure, sensor-derived representations, annotation protocol, and qualitative characteristics of the constructed dataset. Through the inspection of representative samples, the visual criteria adopted to delineate defect boundaries were illustrated across the different sensor channels together with the variability in defect morphology and material appearance present in the collected data. The resulting dataset addresses several limitations identified in the original industrial data by enforcing more consistent acquisition and annotation procedures while explicitly preserving the grouped multi-view structure

of the problem. At the same time, the dataset remains challenging due to severe foreground-background imbalance, large variation in defect scale and topology, and the presence of multiple correlated acquisitions of the same physical defect. These characteristics strongly influence both the construction of the learning pipeline and the definition of the experimental evaluation protocol. The following chapter therefore introduces the evaluation methodology adopted in this work and analyses the implications of the dataset structure for model assessment.

4 Evaluation Methodology and Data Leakage Analysis

The previous chapter described the industrial dataset used throughout this work together with its acquisition protocol and grouped multi-view structure. While repeated acquisitions of the same physical defect are a natural consequence of industrial data scarcity, they also introduce important methodological considerations for model evaluation. Before proceeding to the semantic segmentation methodology, it was therefore necessary to analyse whether the adopted evaluation protocol could provide reliable estimates of model generalization under these dataset conditions. In particular, the presence of correlated acquisitions raises the possibility that standard random splitting procedures may overestimate performance by unintentionally distributing related samples across training and validation subsets. This chapter investigates these issues experimentally and analyses the impact of dataset splitting strategy on evaluation reliability in the considered industrial inspection setting.

4.1 Motivation for evaluation analysis

One of the primary concerns during the initial stages of the project was the limited amount of available industrial defect data. Although repeated multi-view acquisitions increased the total number of available samples, the effective diversity of the dataset remained constrained by the very small number of truly independent physical defect instances. From a supervised learning perspective, this raised an important question regarding the feasibility of obtaining robust generalization under such severe data scarcity conditions. The concern was especially relevant because the target task of this thesis — semantic segmentation — is substantially more demanding than image-level classification and requires reliable pixel-level supervision. At the same time, previously reported classification results obtained on the original industrial dataset showed very high performance despite similarly limited data conditions. Since the dataset had already been used successfully in a classification-oriented setting, these results initially suggested that the inspection problem might still be sufficiently structured for learning-based

approaches to generalize effectively even under severe data scarcity. Consequently, the early stages of the project proceeded under the assumption that, although limited, the available industrial data could still support the development of practical segmentation models.

However, as the dataset structure was analysed in greater detail, it became increasingly evident that the repeated multi-view acquisition protocol illustrated in Fig. 3.4 introduced groups of strongly correlated samples corresponding to different observations of the same physical defect. This observation fundamentally changed the interpretation of the previously reported performance results and raised an important methodological concern: if correlated acquisitions were distributed across both training and validation subsets, the resulting evaluation metrics might not reflect true generalization to unseen defects. The remainder of this chapter investigates this issue experimentally.

4.2 Dataset splitting strategies

A central methodological aspect of the present work concerns the definition of a reliable dataset splitting strategy. As discussed in the previous chapters, the industrial acquisition protocol adopted in this work produces a grouped dataset structure in which multiple images may correspond to different acquisitions of the same physical defect instance. These acquisitions are generated by repositioning or rotating the material under the sensor in order to capture appearance variability induced by orientation and sensing conditions. Consequently, the dataset cannot be treated as a collection of fully independent samples.

Instead, correlations between acquisitions of the same physical defect must be explicitly considered when defining training and validation splits. Evaluating generalization in industrial defect segmentation therefore requires careful consideration of which sources of variability are expected at deployment time. In this work, different splitting strategies are analysed in order to distinguish between generalization to new acquisitions of known defects, generalization to unseen defect instances, and generalization to previously unseen materials.

4.2.1 Random sample-level splitting

The simplest splitting strategy consists of randomly assigning individual images to training and validation subsets independently of their physical origin. Under this protocol, acquisitions corresponding to the same physical defect may appear simultaneously in both subsets. Although the images are not exact duplicates at pixel level, they still share important structural, morphological, and geometric characteristics because they originate from the same underlying defect instance.

This splitting strategy is commonly adopted in standard computer vision settings where

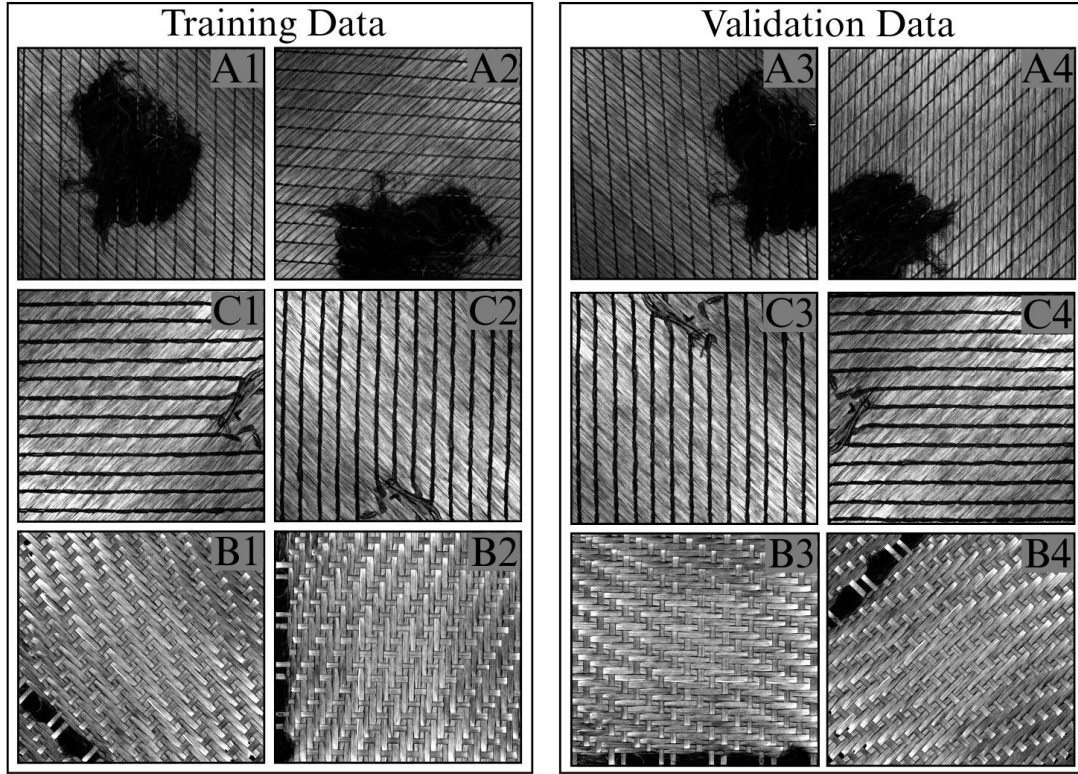


Figure 4.1: Example of a leakage-prone random split. Samples sharing the same letter correspond to different acquisitions of the same physical defect instance obtained under different positions or orientations. Because correlated acquisitions are distributed across both training and validation subsets, the model may partially memorize defect-specific characteristics during training and later recognize related observations during validation.

samples are assumed to be approximately independent. However, this assumption becomes problematic in grouped industrial datasets containing multiple acquisitions of the same physical specimen.

If correlated acquisitions are distributed across both training and validation sets, the model may effectively observe highly related views of the same physical defect during both optimization and evaluation. In such a scenario, the resulting validation metrics may become artificially inflated and may no longer reflect true generalization to unseen defects.

4.2.2 Instance-grouped split

In the instance-grouped split, samples are grouped according to the physical material–defect instance. All acquisitions corresponding to the same defect instance — i.e. multiple positions and orientations of the same physical defect (Fig. 3.4) — are always assigned to the same split. Under this strategy, different defect instances belonging to the same semantic category may appear in both training and validation sets, but no correlated views of the same physical defect

instance are distributed across subsets. This split evaluates the model’s ability to generalize to *new instances of known defect categories* under acquisition variability, while assuming that the underlying material backgrounds observed during training remain representative of those encountered during deployment. Such a scenario is realistic in industrial environments where the set of materials is relatively stable and the primary source of novelty arises from new defect instances rather than from entirely new material domains. Importantly, the grouped acquisition structure itself is not inherently problematic. On the contrary, repeated acquisitions are beneficial because they increase the variability of the available data and expose the model to appearance changes induced by material orientation and sensor interaction. The critical issue is not the presence of correlated acquisitions, but ensuring that such correlations do not cross the training–validation boundary. Although the task considered in this thesis is formulated as binary defect/background segmentation, generalization to new defect instances remains non-trivial. Even without explicit defect-class supervision, defects belonging to the same semantic category may exhibit substantial variability in shape, scale, texture, and local material context. At the same time, this split avoids the additional domain shift introduced by unseen materials, thereby providing a controlled setting for representation analysis, architectural comparison, and hyperparameter studies.

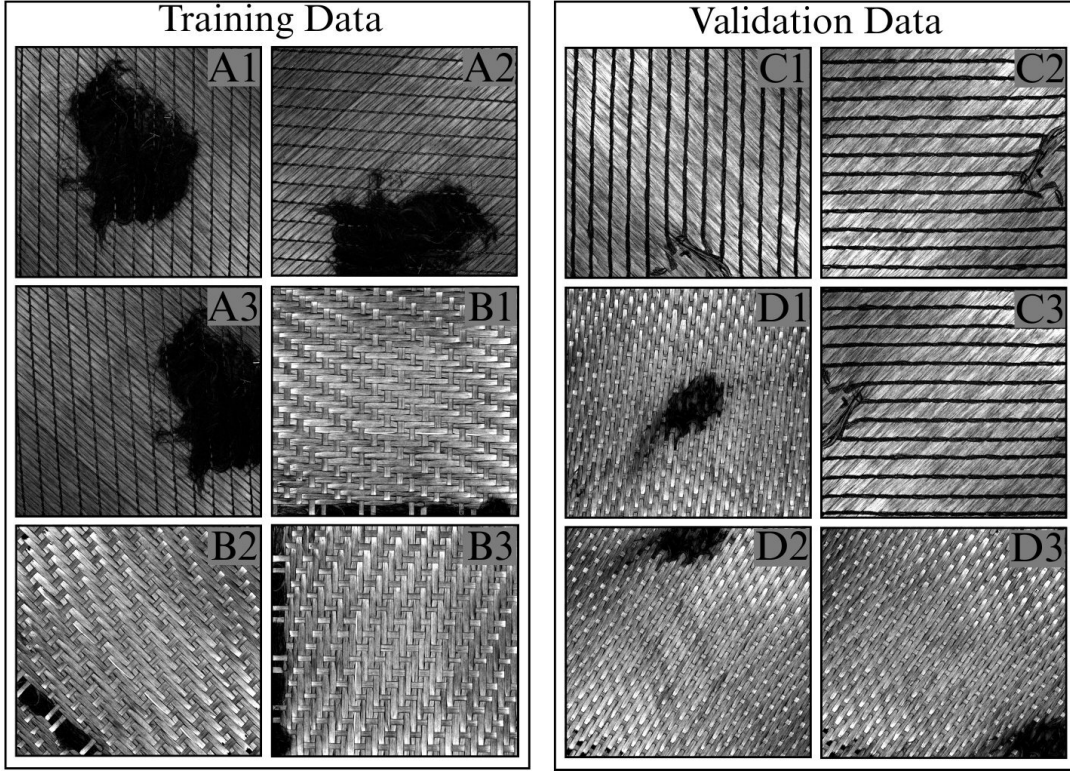


Figure 4.2: Example of a group-consistent split. Samples sharing the same letter correspond to different acquisitions of the same physical defect instance obtained under different positions or orientations. All correlated acquisitions belonging to the same defect group remain confined to a single subset, preventing information leakage across the training–validation boundary.

4.2.3 Material-based split

A second possible splitting strategy is a material-based split, in which all samples belonging to the same material are assigned exclusively either to the training or to the validation set. Material identity is determined by the filename prefix corresponding to the source material specimen. This strategy enforces generalization to *unseen materials* while still evaluating on defect categories that were present during training. Such a scenario is substantially more challenging, since the material itself constitutes the background against which defects must be detected.

In carbon-fibre composites, material-specific texture, weave pattern, fibre orientation, and reflectance behaviour strongly influence the visual manifestation of defects. Consequently, successful performance under such a splitting strategy would indicate that the model has learned defect-related cues that are not tightly coupled to the appearance statistics of specific materials. Due to the limited number of available materials and the resulting reduction in training diversity, this material-generalization regime was not considered sufficiently supported for a reliable experimental evaluation in the present work. Nevertheless, it represents an important direction for future investigation.

	Defect types seen at training	Defect types unseen at training
Material seen at training	(A) Instance generalization New defect instances of <i>known</i> defect types on <i>known</i> materials. Evaluated: instance-grouped split.	(C) New defect-type generalization New defect types (unseen classes) on <i>known</i> materials. Out of scope: insufficient supervision.
Material unseen at training	(B) Material generalization New defect instances of <i>known</i> defect types on <i>unseen</i> materials. Future work: limited material diversity.	(D) Material + new defect-type New defect types on <i>unseen</i> materials. Hardest regime. Open-set scenario beyond the assumptions of supervised segmentation.

Figure 4.3: Conceptual generalization regimes in industrial defect segmentation. Case (A) corresponds to the instance-grouped split evaluated in this work. Case (B) represents generalization to unseen materials and is included as a conceptual future direction. Cases (C) and (D) denote regimes involving previously unseen defect categories.

4.2.4 Conceptual generalization regimes

The experimental evaluation presented in this thesis focuses primarily on the scenario of generalization to new defect instances on known materials, assessed through the instance-grouped split. However, more challenging generalization regimes can also be defined conceptually, as illustrated in Fig. 4.3.

Generalization to previously unseen defect categories is not evaluated in this work. When combined with unseen materials (case D), this setting constitutes an open-set problem for which conventional supervised segmentation is not an appropriate formulation. When restricted to known materials (case C), excluding entire defect categories from training would further reduce the already limited supervision available in the dataset and would prevent reliable learning. For these reasons, the experiments presented in this thesis focus on the scenario in which defect categories are assumed to be known *a priori*, while variability arises from new physical instances of those categories. More challenging generalization settings, particularly those involving unseen materials, are left for future work.

4.3 Reproduction of the classification setup

To analyse the impact of dataset splitting on evaluation reliability, a classification-oriented experimental setup based on the original industrial dataset was reproduced experimentally. The purpose of this analysis was not to exactly reproduce the absolute performance values reported in previous work, nor to optimize classification performance itself. Instead, the objective was to isolate the effect of the splitting strategy while keeping the remaining components of the

experimental pipeline as consistent as reasonably possible. In particular, the reproduced experiments preserved:

- the same industrial dataset;
- the same defect classification task;
- the same Monte Carlo evaluation philosophy;
- and the same overall supervised learning objective.

The principal intentionally modified variable was the dataset splitting logic.

It is important to note that the classification architecture used in the present analysis differs from the one employed in the original classification-oriented study. Rather than reproducing the exact original network architecture, the experiments presented in this thesis employed a SegFormer-based encoder pretrained on ADE20K [9] and adapted for image-level classification through the addition of a classification head.

The precise architectural choice is not the focus of the present analysis. The model is used here primarily as a controlled experimental component for evaluating the effect of different splitting strategies under otherwise comparable training conditions.

4.3.1 Classification setting

The reproduced task consisted of supervised defect classification using the original industrial dataset analysed in Section 3.3. Each sample was assigned a defect category label, and the model was trained to predict the corresponding class from the sensor-derived representations. Although the present thesis focuses primarily on semantic segmentation, the classification setting provides a useful controlled environment for analysing evaluation behaviour under grouped industrial acquisition conditions. In particular, classification offers a simpler and computationally more tractable scenario in which the influence of the splitting methodology can be isolated more clearly before extending the analysis to the substantially more demanding segmentation task.

The reproduced experiments employed sensor-derived input representations consistent with the overall classification-oriented pipeline. Since the objective of the present analysis is to study the relative effect of the splitting methodology rather than the influence of specific representation choices, the detailed representation design is not discussed here and will be addressed separately in Chapter 5.

Because the dataset contains a limited number of independent defect groups, the evaluation results may depend significantly on the particular train–validation partition. For this reason, the reproduced classification experiments relied on repeated randomized grouped evaluations and aggregated performance estimates rather than on a single fixed split. The complete

Monte Carlo validation procedure adopted throughout the experimental analysis is described in Chapter 5.

4.4 Impact of splitting strategy on evaluation

The comparison between random sample-level splitting and group-consistent splitting revealed a substantial difference in evaluation performance.

Under random sample-level splitting, the classification model achieved very high validation metrics, reaching approximately 98.15% accuracy and 96.91% macro F1-score. However, when the splitting protocol was modified to preserve group consistency — ensuring that all acquisitions of the same physical defect remained confined to a single subset — the performance decreased substantially, with accuracy dropping to approximately 76.24% and macro F1-score to approximately 64.49%. Although both metrics exhibit a large degradation, the reduction in macro F1-score is particularly important. The original dataset contains 99 gap defects, 54 fuzzballs, and only 37 missing stitches, resulting in a non-uniform class distribution. Under such conditions, accuracy alone may provide an overly optimistic estimate of performance, since dominant classes contribute more strongly to the final metric. In contrast, macro F1-score evaluates each class independently before averaging the results, thereby providing a more balanced estimate of generalization across defect categories.

To verify whether the same phenomenon also affected semantic segmentation, an analogous comparison was performed using a segmentation model evaluated under the same random and group-consistent splitting protocols. When evaluated using random sample-level splitting, the segmentation model achieved approximately 0.64 IoU, whereas the same model evaluated under instance-grouped splitting achieved approximately 0.43 IoU. Although the absolute metrics differ from those of the classification task, the observed degradation confirms that the inflation effect introduced by correlated acquisitions affects both classification and segmentation performance. Importantly, this substantial performance degradation was observed under otherwise comparable experimental conditions, including the same dataset, classification task, input representation, and overall supervised learning pipeline. The dataset splitting methodology was therefore the principal factor responsible for the observed performance gap.

The effect of correlated acquisitions on evaluation can be understood through a simple analogy. Consider a student preparing for an exam by studying a set of solved questions provided during class. If the final exam contains entirely new questions covering the same concepts, the result reflects the student’s actual understanding and ability to generalize the learned knowledge. However, if the exam instead contains the very same questions seen previously, with only minor variations in presentation, high scores may partially reflect memorization

rather than true understanding.

A similar phenomenon may occur in grouped industrial datasets when different acquisitions of the same physical defect are distributed across both training and validation subsets. Although the validation images are not exact pixel-wise duplicates, they still preserve highly specific structural and morphological characteristics of the same underlying defect instance. Consequently, the resulting validation metrics may substantially overestimate the true generalization capability of the model.

More importantly, random sample-level splitting implicitly evaluates a form of generalization that is only weakly related to the real industrial deployment scenario. Under such a protocol, the model is effectively tested on whether it can recognize alternative acquisitions of defect instances that were already observed during training, typically under slightly different positions or orientations.

In contrast, real industrial inspection systems must generalize to previously unseen physical defects, which may differ substantially in geometry, local context, deformation pattern, reflectance behaviour, and interaction with the surrounding fibre structure. Consequently, very high performance under random sample-level splitting does not necessarily demonstrate robust deployment capability. Instead, it may primarily reflect the model’s ability to recognize highly correlated variations of defect instances already encountered during training.

This issue becomes especially critical under industrial data scarcity, where the number of independent physical defects is extremely limited relative to the total number of acquired images. Under such conditions, the apparent dataset size may become misleading: although repeated multi-view acquisitions increase the number of available samples, the effective diversity of the dataset remains fundamentally constrained by the small number of truly independent defect instances.

4.4.1 Implications for realistic deployment

From an industrial perspective, this distinction is critical. A deployed inspection system must generalize to previously unseen defects encountered during operation rather than simply recognizing alternative views of defects already observed during training. Consequently, evaluation methodologies that ignore the grouped structure of the dataset may lead to unrealistic expectations regarding deployment performance. The analysis presented in this chapter therefore highlights that evaluation methodology is not merely a secondary implementation detail, but a fundamental component of reliable industrial machine learning research.

4.5 Implications of leakage-safe evaluation

The analysis presented in this chapter revealed that random sample-level splitting can substantially inflate performance estimates when multiple acquisitions of the same physical defect instance are present in the dataset. Importantly, this phenomenon was observed in both the classification and semantic segmentation settings. Although the absolute metrics differ between the two tasks, the same general behaviour emerged: random sample-level splitting consistently produced substantially more optimistic results than instance-grouped evaluation. These findings indicate that the observed performance inflation is not a consequence of a specific model architecture or learning task, but rather originates from the splitting strategy itself. When correlated acquisitions of the same physical defect instance are allowed to appear simultaneously in the training and validation subsets, the resulting evaluation no longer measures generalization to previously unseen defects. Instead, it partially rewards the recognition of defect-specific patterns that have already been observed during training.

This observation has important practical implications for industrial inspection systems. Under real deployment conditions, the objective is not to recognize additional views of defects that have already been encountered during training, but rather to correctly identify new defect instances appearing on previously unseen material samples. Consequently, evaluation protocols that permit information leakage between training and validation data may substantially overestimate the performance that can realistically be expected in production environments. The identification of this issue fundamentally changes the interpretation of the segmentation problem investigated in the remainder of this thesis. Once the influence of correlated acquisitions is removed, the true scarcity of independent defect instances becomes much more apparent, revealing a substantially more challenging and realistic industrial learning scenario.

These observations directly motivate the experimental methodology adopted throughout the remainder of this work. In particular, they justify:

- the adoption of strict instance-grouped splitting protocols;
- the investigation of augmentation strategies designed to increase effective training diversity;
- the study of representation choices capable of extracting maximal information from limited data;
- and the use of repeated Monte Carlo evaluation to reduce split-dependent variability.

Although leakage-safe evaluation produces substantially lower numerical metrics, it provides a significantly more realistic estimate of model generalization under practical industrial conditions. The results of this chapter therefore establish the methodological foundation for

the remainder of the thesis. All subsequent segmentation experiments adopt the instance-grouped leakage-safe evaluation protocol introduced in this chapter.

5 Experimental and Methodological Framework

The analysis presented in Chapter 4 showed that evaluation under grouped industrial acquisition conditions is substantially more challenging than initially expected. In particular, once correlated acquisitions of the same physical defect are prevented from appearing simultaneously in training and validation sets, model performance decreases significantly. This observation indicates that the scarcity of truly independent defect instances constitutes one of the central limitations of the considered inspection problem. At the same time, this result does not imply that further methodological investigation becomes meaningless. Rather, it changes the interpretation and objectives of the subsequent analysis. The goal is no longer simply to maximize performance metrics under potentially optimistic evaluation protocols, but instead to investigate which methodological choices remain beneficial when realistic leakage-safe evaluation strategies are enforced.

Under severe industrial data scarcity, several aspects of the learning pipeline become especially important. The limited number of independent training examples increases the importance of the input representation, the compatibility between sensor-derived data and pretrained architectures, the robustness of augmentation strategies, and the overall training methodology. These factors may substantially influence the model’s ability to generalize beyond the specific defect instances observed during training.

For this reason, the remainder of this thesis focuses on a systematic investigation of robustness-oriented methodological choices for supervised semantic segmentation under constrained industrial conditions. In particular, the following aspects are analysed:

- representation and normalization strategies for the sensor-derived channels produced by the Fscan system;
- compatibility between industrial multi-channel representations and RGB-pretrained architectures;
- augmentation strategies designed to increase effective training variability under limited supervision;

- the influence of model capacity and optimization choices;
- training methodologies compatible with leakage-safe evaluation protocols.

The objective of these investigations is not to claim that the scarcity problem is fully solved. Instead, the aim is to understand which methodological choices remain most effective under realistic industrial constraints, to identify the limitations that persist despite these strategies, and to obtain a more reliable picture of the practical feasibility of supervised segmentation for carbon-fibre defect inspection.

The chapter is organized as follows. First, the characteristics of the sensor-derived representations and the adopted normalization strategies are discussed. Subsequently, the segmentation architectures, augmentation techniques, training procedures, and evaluation methodologies used throughout the experimental analysis are presented.

5.1 Input representation design

The industrial dataset introduced in Chapter 3 provides a set of derived feature channels extracted from the Fscan acquisition system. These channels encode complementary physical properties of the carbon-fibre surface and therefore differ fundamentally from the RGB images on which most modern vision architectures are pretrained. This section focuses on how the sensor-derived channels are selected, combined, and transformed when used as input to the segmentation models. Particular attention is given to the mismatch between physically-derived feature channels and models pretrained on natural RGB images, and to the implications that this mismatch introduces for representation design choices. The representation pipeline introduced in this chapter is used consistently throughout the experimental analysis presented in Chapter 6.

5.1.1 Human versus neural network interpretability

From a human perspective, the derived feature channels previously shown (Fig. 3.3) are often more informative than the raw images shown in (Fig. 3.2): they suppress illumination artefacts and enhance visually salient defect patterns. However, this does not imply that the same channels are automatically optimal for a neural network. A neural network does not reason semantically; it processes numerical values arranged on a spatial grid and learns statistical regularities through optimization.

A channel is therefore useful only insofar as:

- its numerical encoding is compatible with the model architecture and pretraining assumptions;

- its dynamic range and distribution do not introduce artificial biases;
- it does not contain discontinuities that can be misinterpreted as meaningful edges.

This distinction is particularly important when using models pretrained on RGB natural images, such as SegFormer. The derived channels are obtained through photometric stereo analysis, where multiple illumination directions are used to estimate per-pixel surface orientation and reflectance properties [10].

5.1.2 Channel complementarity and representation selection

The industrial dataset provides five derived feature channels, whereas most pretrained segmentation architectures expect three-channel inputs. This mismatch introduces a representation design problem, since different channels emphasize complementary physical properties of the carbon-fibre surface, including fibre orientation and reflectance behaviour under directional illumination. Visual inspection of the dataset revealed that the usefulness of each sensor channel strongly depends on both the material characteristics and the defect morphology. No single representation was found to be consistently informative across all defect categories and acquisition conditions. Figure 5.1 illustrates two representative examples. In the first case, corresponding to a gap defect, the azimuthal channel clearly highlights the defect boundary through strong orientation discontinuities, while the diffuse and fibular representations provide substantially weaker structural contrast. Conversely, in the second example, corresponding to a missing stitch defect, the azimuthal representation provides limited discriminative information, whereas the reflectance-based channels more clearly reveal the defective region.

These observations suggest that different channels provide complementary information depending on the material properties and the defect morphology. This introduces a representation design problem. On one hand, restricting the input to three channels preserves direct compatibility with pretrained segmentation architectures originally designed for RGB imagery. On the other hand, modifying the input projection layer allows the model to exploit the full set of sensor-derived representations at the cost of partially reducing compatibility with pretrained weights.

Two principal strategies were therefore investigated.

Channel subset selection. A subset of three channels can be selected and directly mapped to the pretrained input projection layer of the segmentation architecture. This strategy preserves full architectural compatibility with standard pretrained models and maximizes reuse of weights learned from large-scale natural-image datasets. At the same time, restricting the input representation to only three channels introduces an implicit information selection process, since certain physical properties captured by the sensor are discarded. Consequently, the

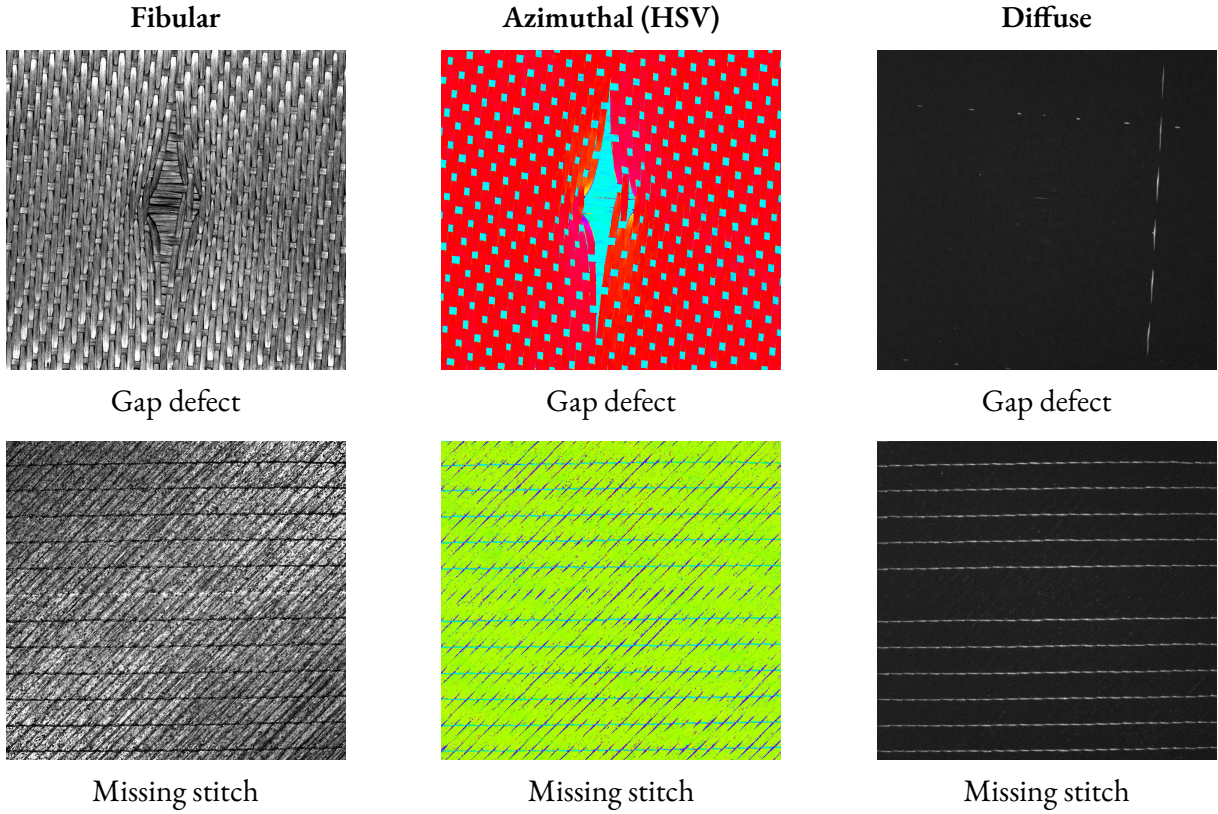


Figure 5.1: Comparison of different sensor representations for two defect instances. Columns correspond to the fibular, azimuthal, and diffuse channels, respectively. For visualization purposes, the azimuthal channel is rendered using an HSV colormap to emphasize orientation variations; however, the underlying representation used during training remains the original single-channel orientation tensor.

effectiveness of this strategy depends strongly on which channels are retained and on how informative they are for the considered material and defect configurations.

Based on the qualitative analysis of the sensor representations, the azimuthal, fibular, and diffuse channels were selected as the primary candidates for the three-channel configuration. The azimuthal representation provides orientation-related information, while the diffuse and fibular channels capture complementary reflectance characteristics. In particular, the fibular representation often acts as an intermediate representation between orientation-based and reflectance-based cues, providing relatively stable structural contrast across multiple defect categories.

The channel-selection problem therefore represents a trade-off between maximizing compatibility with pretrained representations and preserving the diversity of complementary information available in the multi-channel sensor data.

Input projection modification. An alternative strategy consists of modifying the input projection layer of the segmentation architecture to directly accept all five sensor channels. In principle, this approach preserves the full amount of available sensor information and allows the model to jointly exploit multiple complementary representations of the material surface. However, most pretrained segmentation architectures are designed for three-channel RGB inputs and therefore rely on initialization statistics learned from natural-image distributions.

Extending the input layer to five channels partially reduces the benefit of pretrained initialization, since the additional input dimensions cannot directly exploit the pretrained weights in their original form. Furthermore, the higher-dimensional input space introduces additional trainable parameters in the first projection layer and may increase the difficulty of learning robust representations under severe data scarcity conditions.

Consequently, a trade-off emerges between preserving the full sensor information and maintaining effective transfer-learning behaviour from pretrained models. These alternatives define a representation design space explored experimentally in Chapter 6.

5.1.3 Frequency-based derived representations

Additional frequency-domain transformations were investigated in order to exploit the highly regular spatial structure of woven carbon-fibre textures. Carbon-fibre materials exhibit periodic arrangements of fibre tows, which generate strong directional components in the spatial-frequency domain. Defects often appear as local disruptions of this periodic organization. For this reason, frequency-domain filtering was explored as a possible strategy for separating large-scale structural organization from local texture irregularities.

Fourier-domain analysis

The two-dimensional Fast Fourier Transform (FFT) provides a representation of the image in terms of its spatial-frequency components. Low spatial frequencies correspond to slowly varying large-scale structures, whereas high spatial frequencies capture fine-grained texture details and local spatial variations. In the present work, the frequency-domain transformations were applied primarily to the diffuse representation, since qualitative inspection suggested that this channel preserved the most stable large-scale structural information across materials. Applying the FFT to carbon-fibre images often produces characteristic star-shaped spectra reflecting the strong directional periodicity of the fibre layout. Figure 5.2 shows a representative example.

Low-pass and high-pass filtering

Two principal filtering strategies were considered:

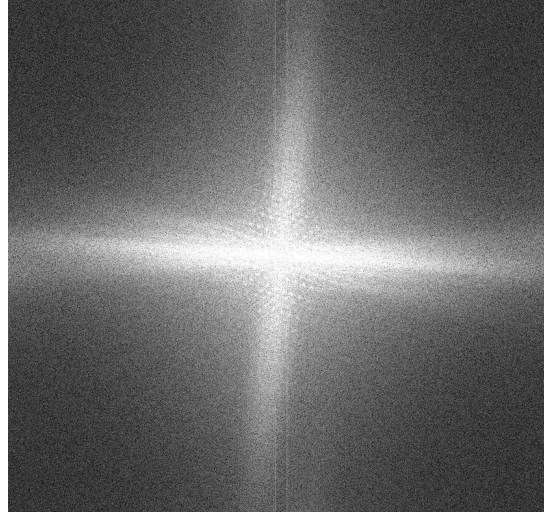


Figure 5.2: Example of the frequency spectrum of a carbon-fibre surface obtained through the Fast Fourier Transform.

- low-pass filtering, emphasizing large-scale structural organization;
- high-pass filtering, emphasizing local texture irregularities.

Low-pass filtering suppresses high-frequency components while preserving the broader structural organization of the textile. Conversely, high-pass filtering removes low-frequency information and emphasizes local texture variations and fine structural discontinuities. In practice, the high-pass representation was obtained by subtracting the low-pass filtered image from the original image:

$$I_{HP} = I - I_{LP}$$

where I denotes the original image and I_{LP} the low-pass filtered representation. Representative examples are shown in Figure 5.3.

Visual inspection suggested that, for certain defect types and material configurations, low-pass filtering of the diffuse channel may improve the visibility of broader structural discontinuities by suppressing high-frequency textile texture. In particular, reducing fine-grained fibre patterns can sometimes make large-scale irregularities associated with defect regions more visually apparent.

However, this behaviour was not observed consistently across all materials and defect categories. In several cases, high-frequency details remained important for preserving defect visibility and boundary definition.

For this reason, frequency-based representations were treated as an additional representation-design alternative rather than as a universally beneficial transformation. The following input configurations were therefore selected for experimental evaluation:

- azimuthal + fibular + low-pass filtered diffuse;

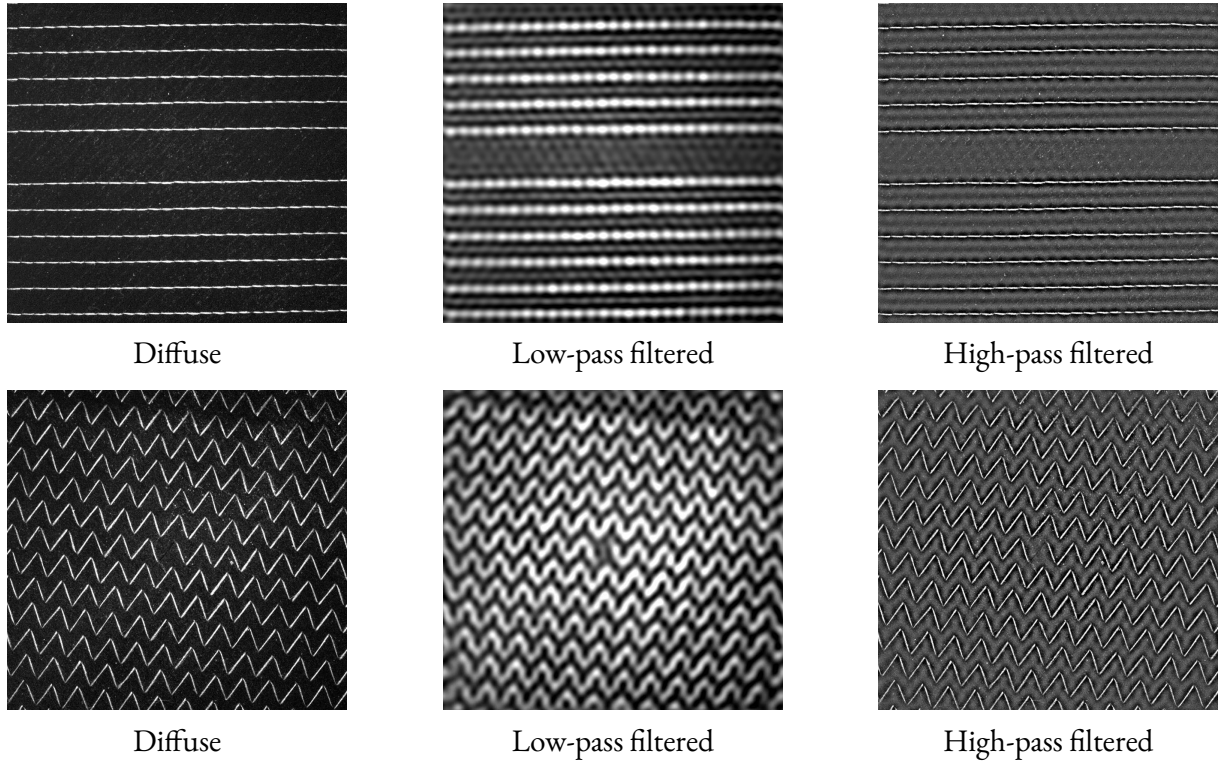


Figure 5.3: Examples of frequency-based representations applied to diffuse-channel images.

- azimuthal + fibular + high-pass filtered diffuse.

FFT filtering radius

For frequency-based representations, the radius of the Fourier-domain low-pass filter determines the amount of high-frequency information retained in the transformed image. Small radii emphasize global textile organization while suppressing local texture details, whereas larger radii preserve more high-frequency structure. Several radius values were therefore investigated experimentally in order to identify the most suitable trade-off between structural smoothing and detail preservation.

5.2 Normalization and pretrained compatibility

After defining the sensor-derived input representations, an additional preprocessing stage is required to ensure numerical compatibility with pretrained segmentation backbones and to reduce inter-scan variability across acquisitions. Several normalization challenges arise in this setting.

First, some channels encode angular quantities rather than intensity values, introducing scale and representation mismatches with standard RGB-pretrained architectures. Second, the dy-

dynamic range of the sensor channels may vary substantially across scans due to material properties, fibre orientation, and acquisition conditions. Finally, the pretrained SegFormer encoder expects input statistics compatible with those encountered during large-scale pretraining.

The normalization strategies presented in this section were therefore designed to stabilize the input representations while preserving the structural information relevant for defect localization.

5.2.1 Azimuthal channel representation

The azimuthal channel poses two fundamental challenges. First, azimuthal values represent angular quantities rather than light intensities, leading to a scale mismatch with both the other sensor channels and the statistics expected by RGB-pretrained models. Second, angular variables are inherently circular, which may introduce artificial discontinuities in the spatial representation of the fibre orientations.

Scale mismatch. SegFormer and similar architectures are pretrained on ImageNet [11], where input values represent light intensities and follow specific normalization statistics. Feeding azimuthal values directly in the range $[-\pi, +\pi]$ forces the network to process angular quantities as if they were standard intensity values and introduces a strong scale mismatch with the other feature channels, which lie approximately in $[0,1]$.

Circularity and artificial discontinuities. Angles are circular variables. Two neighbouring pixels with values close to $+\pi$ and $-\pi$ represent nearly identical fibre orientations, yet their numerical difference in Euclidean space is approximately 2π . This behaviour originates from the periodic nature of angular representations: the values $-\pi$ and $+\pi$ correspond to the same angular boundary and should therefore be considered adjacent rather than maximally distant. However, when angular values are represented as standard scalar intensities, this periodic structure is lost.

As a consequence, smooth orientation transitions in the physical material may appear as abrupt discontinuities in the image representation. Early convolutional filters operate through local differences between neighbouring pixel values and may therefore interpret these artificial transitions as strong edges or meaningful texture boundaries, despite having no physical counterpart in the underlying fibre structure.

This issue is particularly relevant for highly regular carbon-fibre textures, where orientation continuity represents an important structural cue.

Directional angular mapping. To address the scale mismatch, an initial normalization strategy maps the azimuthal channel from $[-\pi, +\pi]$ to $[0,1]$ using a linear transformation:

$$f_{\text{dir}}(\theta) = \frac{\theta + \pi}{2\pi}.$$

This mapping preserves the full angular direction and ensures numerical compatibility with the other channels. However, it does not resolve the circularity issue. As a numerical example, consider

$$\theta_1 = -\pi + 0.01, \quad \theta_2 = +\pi - 0.01.$$

Although these angles correspond to nearly identical fibre orientations, the normalized values become

$$f_{\text{dir}}(\theta_1) \approx 0.0016, \quad f_{\text{dir}}(\theta_2) \approx 0.9984,$$

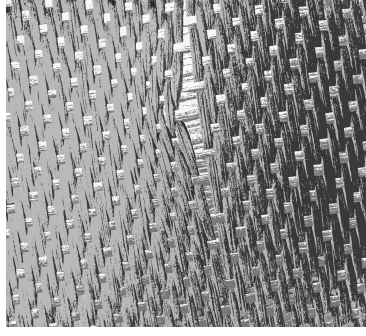
thereby introducing an artificial high-contrast transition that has no physical counterpart.

Orientation-invariant mapping. Carbon-fibre tows are undirected structures: a fibre oriented at angle θ is physically equivalent to one oriented at $\theta + \pi$. To reflect this property, an alternative normalization collapses opposite directions into a single orientation by applying a modulo- π operation:

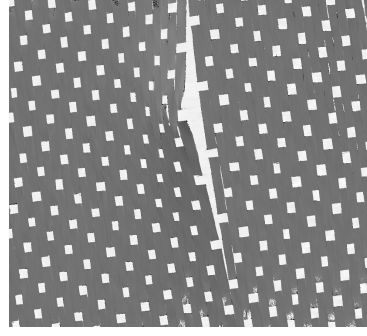
$$\theta_{\text{mod}} = (\theta + \pi) \bmod \pi, \quad f_{\text{ori}}(\theta) = \frac{\theta_{\text{mod}}}{\pi}.$$

Under this mapping, angles differing by π are assigned the same normalized value. This representation removes directional ambiguity and reduces spurious discontinuities, resulting in a smoother spatial encoding that better matches the underlying material structure. A visual comparison between the directional and orientation-invariant mappings is shown in Fig. 5.4.

Implications. The orientation-only mapping reduces angular discontinuities and produces a smoother representation of fibre orientation. However, it also removes directional information that may still contain useful cues for defect localization. The relative usefulness of the two mappings therefore remains an empirical question and is investigated experimentally in Chapter 6.



(a) Directional normalization



(b) Orientation-only normalization

Figure 5.4: Visualization of two normalization strategies for the azimuthal channel. Directional normalization preserves angular direction but creates artificial discontinuities at the $-\pi/+ \pi$ boundary. Orientation-only normalization applies a modulo- π mapping that merges opposite directions and yields a smoother representation of fibre orientation.

5.2.2 Per-sample normalization

After defining the channel mappings and derived representations, an additional normalization stage is required to reduce inter-scan variability and improve compatibility with pre-trained architectures. A per-sample, per-channel min-max normalization strategy was therefore adopted. The objective is not to preserve absolute physical magnitudes, but rather to emphasize relative spatial structure within each scan while maintaining numerical compatibility with pretrained encoders.

After the channel-specific transformations described previously, normalization is applied independently to each selected channel:

$$\tilde{c}(x) = \frac{c(x) - \min_x c(x)}{\max_x c(x) - \min_x c(x) + \varepsilon}.$$

Although several channels are already nominally bounded in the interval $[0,1]$, their effective dynamic range within an individual scan is often substantially narrower due to fibre alignment, illumination conditions, and field-of-view variability. Under these conditions, per-sample normalization effectively acts as a form of contrast stretching, amplifying relative spatial variations that may be informative for defect localization.

As a consequence, the normalized representations primarily encode relative spatial deviations within the current sample rather than absolute global intensity levels. This discourages the model from exploiting scan-specific global offsets or material-dependent intensity statistics as shortcut cues, and instead encourages learning from localized structural irregularities associated with defect regions.

5.2.3 Compatibility with RGB-pretrained backbones

After per-sample normalization, standard ImageNet mean/std normalization is applied before feeding the inputs to the SegFormer encoder. Although the sensor-derived representations differ substantially from natural RGB imagery, this strategy was retained for compatibility with the RGB-pretrained backbone and was evaluated empirically in the experimental analysis.

The SegFormer encoder used in this work originates from ImageNet-1K pretraining and is subsequently fine-tuned for semantic segmentation on ADE20K. Although the encoder statistics are therefore no longer identical to those of the original ImageNet classification setting, the pretrained weights still implicitly assume inputs processed under the same normalization convention used during pretraining.

The purpose of ImageNet normalization in this context is not to match the semantic content or physical interpretation of RGB images, since the sensor-derived channels used in this work differ substantially from natural-image color channels. Instead, the normalization is intended to maintain numerical compatibility with the activation ranges expected by the pre-trained encoder during initialization and early optimization.

The two normalization stages therefore serve complementary purposes. Per-sample min-max normalization reduces inter-scan variability and emphasizes relative spatial structure within each sample, whereas ImageNet normalization improves compatibility with the pre-trained encoder initialization.

This strategy was adopted only for architectures initialized from ImageNet-pretrained parameters. Its influence on optimization stability and segmentation performance is analysed experimentally in Chapter 6.

5.3 Segmentation architecture and transfer-learning strategy

This section presents the segmentation architectures and transfer-learning strategies adopted throughout the experimental analysis. Particular attention is given to the challenges introduced by severe industrial data scarcity, the non-standard nature of the sensor-derived representations, and the compatibility constraints associated with RGB-pretrained segmentation backbones. The section also discusses the adaptation of pretrained semantic segmentation models to the binary industrial defect-segmentation setting considered in this thesis.

5.3.1 SegFormer architecture

The primary segmentation architecture adopted in this work is SegFormer, a transformer-based semantic segmentation framework [1] composed of a hierarchical encoder and a lightweight

all-MLP decoder.

SegFormer was selected for several reasons particularly relevant to the industrial setting considered in this thesis.

First, the hierarchical Transformer encoder allows the model to capture both fine-grained local texture irregularities and larger structural patterns across the textile surface. This property is important for carbon-fibre inspection, where defects may appear either as small local disruptions or as larger structural distortions extending across multiple fibre bundles. Furthermore, the architecture provides a favourable trade-off between representational capacity, memory consumption, and computational efficiency, making it suitable for extensive Monte Carlo experimentation under limited computational resources.

5.3.2 Pretrained checkpoint adaptation

The pretrained SegFormer checkpoints used in this work follow a two-stage training procedure. First, the hierarchical encoder is pretrained on the ImageNet-1K classification dataset. Subsequently, the full segmentation model is fine-tuned on semantic segmentation datasets such as ADE20K.

The pretrained checkpoints adopted in this work correspond to the publicly available SegFormer models fine-tuned on ADE20K semantic segmentation available through the HuggingFace model repository. Although alternative checkpoints fine-tuned on datasets such as Cityscapes are also available, investigating the influence of the source segmentation dataset lies beyond the scope of the present work.

The pretrained ADE20K segmentation models are designed for multi-class semantic segmentation and therefore employ an output projection layer producing 150 semantic classes. In the present work, however, the task is formulated as binary defect segmentation, where the model predicts a single foreground defect mask.

Consequently, the original segmentation head is not directly compatible with the target task and must be replaced by a newly initialized binary segmentation head. Only the encoder weights can therefore be reused directly from the pretrained checkpoint.

This limitation is not particularly restrictive in practice because the SegFormer decoder is intentionally lightweight relative to the encoder. Most transferable visual representations are therefore expected to reside in the pretrained encoder stages rather than in the final segmentation head.

5.3.3 Backbone capacity variants

Several SegFormer backbone variants with progressively increasing encoder capacity were evaluated experimentally:

- SegFormer-B0;
- SegFormer-B2;
- SegFormer-B5.

According to the original SegFormer architecture specification, the considered encoder variants differ substantially in model capacity. The MiT-B0 encoder contains approximately 3.4 million parameters, whereas MiT-B2 contains approximately 24.2 million parameters. In contrast, the large-scale MiT-B5 encoder increases the encoder capacity to approximately 81.4 million parameters. The decoder remains comparatively lightweight, containing approximately 0.4 million parameters for B0 and approximately 3.3 million parameters for the larger variants. Increasing backbone size provides higher representational capacity but also substantially increases the number of trainable parameters. Under severe industrial data scarcity, excessively large models may therefore become more prone to overfitting rather than improved generalization. The influence of encoder capacity on segmentation performance and generalization behaviour is analysed experimentally in Chapter 6.

5.4 Optimization under severe data scarcity

The limited number of independent defect instances available for training introduces significant optimization challenges for semantic segmentation. In particular, severe foreground–background imbalance and the risk of overfitting require optimization strategies specifically designed to promote stable learning and robust generalization under limited data conditions. This section presents the loss formulations, learning-rate strategies, and regularization mechanisms adopted throughout the experimental analysis.

5.4.1 Focal-Tversky loss formulation

Previous work on industrial fabric defect segmentation [4] employed a focal overlap-based loss formulation in order to address the severe foreground–background imbalance typical of textile inspection tasks. This observation motivated the use of a focal overlap-based objective in the present work as well. Rather than adopting the focal–Dice formulation used by Xu et al., this thesis employs the Focal–Tversky loss [12], which combines focal modulation with the

Tversky overlap formulation originally introduced for highly imbalanced segmentation problems [13]. The principal motivation for this choice is that the Tversky formulation allows asymmetric penalization of false positives and false negatives through dedicated weighting parameters. This asymmetric behaviour is particularly important in industrial inspection settings, where failing to detect a defect is generally more costly than producing occasional false alarms. Consequently, the Focal–Tversky formulation provides a natural mechanism for emphasizing defect recall while simultaneously addressing the severe class imbalance present in the dataset. The Tversky index is defined as [13]:

$$TI = \frac{TP}{TP + \alpha FP + \beta FN}$$

where TP , FP , and FN denote the number of true positive, false positive, and false negative pixels respectively.

The parameters α and β control the relative penalization of false positives and false negatives. When $\beta > \alpha$, false negatives receive a stronger penalty, encouraging the model to prioritize defect detection recall.

The Focal–Tversky formulation further emphasizes difficult examples through an additional focusing exponent [12]:

$$L_{FT} = (1 - TI)^\gamma$$

where γ controls the strength of the focal modulation.

When $\gamma > 1$, the contribution of well-segmented regions with high Tversky index is progressively suppressed, whereas poorly segmented regions contribute proportionally more to the loss. Consequently, increasing γ encourages the optimization process to focus more strongly on difficult examples, hard boundaries, and small defect regions that may otherwise be dominated by the large background class.

This formulation therefore combines two desirable properties:

- asymmetric penalization of false negatives through the Tversky formulation;
- increased emphasis on difficult or poorly segmented regions through the focal modulation term.

The Focal–Tversky loss is particularly suitable for industrial defect segmentation under severe class imbalance, where defect regions are small and subtle texture anomalies may otherwise be overwhelmed by the dominant background class.

In the experiments presented in this thesis, the majority of the training configurations adopt a Focal–Tversky formulation with $\alpha = 0.2$ and $\beta = 0.8$, thereby assigning a stronger

penalty to missed defect regions. The focal modulation exponent was set to $\gamma = 1.75$. This value provides substantial emphasis on difficult and uncertain examples while remaining slightly less aggressive than the commonly used default value $\gamma = 2$, thereby promoting more stable optimization under severe data scarcity conditions.

5.4.2 Learning-rate strategies

Several learning-rate configurations were investigated experimentally, since optimization behaviour can significantly affect generalization under severe data scarcity conditions. In addition to constant learning rates, alternative scheduling strategies were explored, including:

- cosine annealing schedules [14];
- OneCycle learning-rate policies [15];
- separate learning rates for encoder and decoder components.

The SegFormer encoder is initialized from pretrained weights, whereas the segmentation head is initialized randomly. Consequently, the two components may exhibit different optimization dynamics during training. This motivates the investigation of separate learning rates, where the pretrained encoder can be updated more conservatively while allowing faster adaptation of the randomly initialized decoder.

At the same time, the optimal optimization strategy under severe data scarcity is not obvious a priori. Simpler configurations based on a single global learning rate may also provide stronger regularization and more stable convergence behaviour. For this reason, multiple learning-rate strategies were evaluated experimentally rather than assuming a preferred configuration in advance.

Several of these optimization strategies were explored during the experimental development phase in order to assess their influence on convergence stability and generalization behaviour under limited data conditions. However, since the observed differences were generally less significant than those associated with representation design and leakage-safe evaluation, only the most relevant configurations are discussed in detail in the experimental chapter.

5.4.3 DropPath regularization

DropPath is a stochastic regularization strategy commonly adopted in modern transformer-based architectures. During training, residual branches within the network are randomly dropped with a specified probability, forcing the model to avoid excessive dependence on individual computation paths and encouraging more distributed feature representations.

This mechanism is closely related to stochastic depth regularization. Instead of removing individual activations as in standard dropout, DropPath suppresses entire residual transformations during training. As a consequence, different forward passes effectively traverse slightly different subnetworks, which may improve robustness and generalization.

In hierarchical transformer architectures such as SegFormer, DropPath regularizes the propagation and fusion of multi-scale features across encoder stages. This is particularly relevant under severe industrial data scarcity, where large transformer models may otherwise overfit to highly specific texture configurations or defect appearances present in the training set.

The DropPath probability controls the strength of the regularization. Higher probabilities increase the stochasticity of the optimization process and encourage stronger regularization, whereas excessively large values may impair convergence stability under limited-data conditions.

The default DropPath probability used in the original SegFormer implementation is 0.1. Since the optimal regularization strength under severe industrial data scarcity is not obvious a priori, multiple DropPath configurations were explored during the experimental development phase in order to assess their influence on convergence stability and generalization behaviour.

5.5 Leakage-safe evaluation methodology

The experiments presented in this thesis were designed to minimize the influence of dataset leakage and split-dependent variability. Particular care was devoted to ensuring fair comparisons between different configurations despite the limited number of available defect instances.

All experiments adopt the leakage-safe evaluation methodology introduced and analysed in Chapter 4. In particular, unless stated otherwise, the reported experiments use the instance-grouped splitting strategy described in Section 4.2.2, where all acquisitions associated with the same physical defect instance are assigned exclusively either to the training or validation split. This protocol is adopted in order to evaluate generalization to previously unseen defect instances while preventing information leakage caused by correlated acquisitions.

5.5.1 Monte Carlo evaluation protocol

Due to the limited size of the dataset, model performance can vary significantly depending on the specific training-validation split. To mitigate this effect, a Monte Carlo evaluation strategy was adopted. This is particularly important under severe industrial data scarcity, where the small number of independent defect instances may otherwise cause performance estimates to depend excessively on a specific train-validation partition. For each experiment, multiple dataset splits were generated using different random seeds. The train-validation proportion

was kept fixed across all runs in order to maintain comparable training conditions while varying only the specific allocation of defect instances between subsets. Each configuration was therefore trained and evaluated multiple times, and the reported results correspond to the mean and standard deviation across the runs. Most experiments were evaluated using ten seeds. Some intermediate hyperparameter studies used three seeds sampled from the same set of ten in order to reduce the computational cost while preserving partial comparability between configurations. Since the same subset of seeds was reused consistently across these intermediate experiments, the resulting comparisons remain internally coherent even if they are not directly comparable with the full ten-seed evaluations.

5.5.2 Experimental setup

All experiments were conducted using the PyTorch deep-learning framework on an NVIDIA A100 GPU. Particular care was devoted to ensuring fair and reproducible comparisons across the large number of experimental configurations investigated throughout this thesis. Since the objective of the experimental analysis is to isolate the effect of representation choices, normalization strategies, architectural variants, and optimization configurations, it is important to minimize variability originating from inconsistent data partitions.

For this reason, the dataset splits generated for each Monte Carlo seed were stored explicitly and reused across all experiments. Consequently, different configurations evaluated under the same seed were always trained and validated on exactly the same acquisition groups.

This procedure ensures that performance differences between configurations are primarily attributable to the investigated methodological changes rather than to accidental variations in the train-validation partition.

In addition, all experiments were performed under the leakage-safe group-consistent splitting protocol described previously, thereby preventing correlated acquisitions of the same physical defect instance from appearing simultaneously in both training and validation subsets.

5.5.3 Evaluation metrics

The performance of the segmentation models is evaluated using standard pixel-wise metrics commonly adopted in semantic segmentation.

Intersection-over-Union metrics

The primary evaluation metric adopted in this work is the Intersection-over-Union (IoU), defined as:

$$IoU = \frac{TP}{TP + FP + FN}$$

where TP , FP , and FN denote the number of true positive, false positive, and false negative pixels respectively. The standard formulation above corresponds to the *hard IoU*, where the predicted probability map is first converted into a binary segmentation mask through thresholding. In practice, a threshold value of 0.5 is commonly adopted in the literature for binary segmentation tasks. However, this threshold is not necessarily optimal in highly imbalanced industrial inspection scenarios such as the one considered in this thesis, where defect pixels represent only a very small fraction of the image. A fixed threshold may artificially penalize predictions that correctly localize defect regions with moderate confidence values while failing to exceed the selected cutoff. Although both hard IoU and soft IoU were monitored during the experiments, the majority of the quantitative comparisons reported in Chapter 6 focus on hard IoU, since it more directly reflects the quality of the final thresholded segmentation masks used in practical inspection scenarios. Unless stated otherwise, the reported IoU values correspond to a global (**micro**) formulation computed over the aggregated pixel counts of the validation set rather than to a per-image averaged IoU. This choice was motivated by the strong variability in defect size across samples. Under a purely per-image averaging scheme, failing to segment a very small defect instance would contribute equally to the final metric as missing a substantially larger defect region. By aggregating pixel statistics globally, larger defect regions contribute proportionally more to the final metric, resulting in a measure that better reflects the overall segmentation quality at the pixel level.

The soft formulation is defined as:

$$IoU_{\text{soft}} = \frac{\sum_i p_i g_i}{\sum_i p_i + \sum_i g_i - \sum_i p_i g_i}$$

where $p_i \in [0,1]$ denotes the predicted probability for pixel i and $g_i \in \{0,1\}$ represents the corresponding ground-truth label. Unlike hard IoU, the soft formulation does not require thresholding and therefore evaluates the degree of overlap directly at the probability level. This provides a smoother and often more informative measure of segmentation quality, particularly in situations where the model produces uncertain predictions near defect boundaries.

Precision and recall

Precision and recall are defined as:

$$Precision = \frac{TP}{TP + FP}.$$

$$Recall = \frac{TP}{TP + FN}.$$

In industrial inspection scenarios, missing a defect is typically more costly than raising a false alarm. For this reason, recall plays an important role in the interpretation of the results.

5.6 Data augmentation via CutMix

The limited number of independent defect instances available in the dataset makes data augmentation an important component of the training procedure. In addition to the geometric variability already introduced during acquisition through controlled repositioning of the material samples under the sensing system, further image-level augmentation was investigated during training. In particular, a more structured augmentation strategy based on *CutMix* [16] was explored. Originally proposed for image classification, CutMix generates new training samples by combining regions extracted from different images. In the context of semantic segmentation, this idea can be adapted by transferring annotated defect regions from one image onto another background while preserving the corresponding segmentation mask.

CutMix generation procedure

In this work, CutMix samples were generated online during training. For each augmented sample, a defect region was extracted from a donor image using the available segmentation annotation and pasted onto a defect-free material texture. Let I_d denote a donor image containing a defect and M_d the corresponding defect mask. A new training sample I_{mix} is generated as:

$$I_{mix} = M_d \odot I_d + (1 - M_d) \odot I_b$$

where I_b denotes the clean background image and $\odot$ represents element-wise multiplication. The background image (I_b) was sampled from a separate auxiliary pool of non-defective material scans. This pool was not part of the annotated defect dataset described in Chapter 3 and was not included in the dataset statistics, since it did not contain defect annotations. Its role was limited to providing clean carbon-fibre texture backgrounds for synthetic CutMix generation. The segmentation mask associated with the generated sample is directly obtained from the donor mask M_d . The generation procedure is illustrated visually in Fig. 5.5.

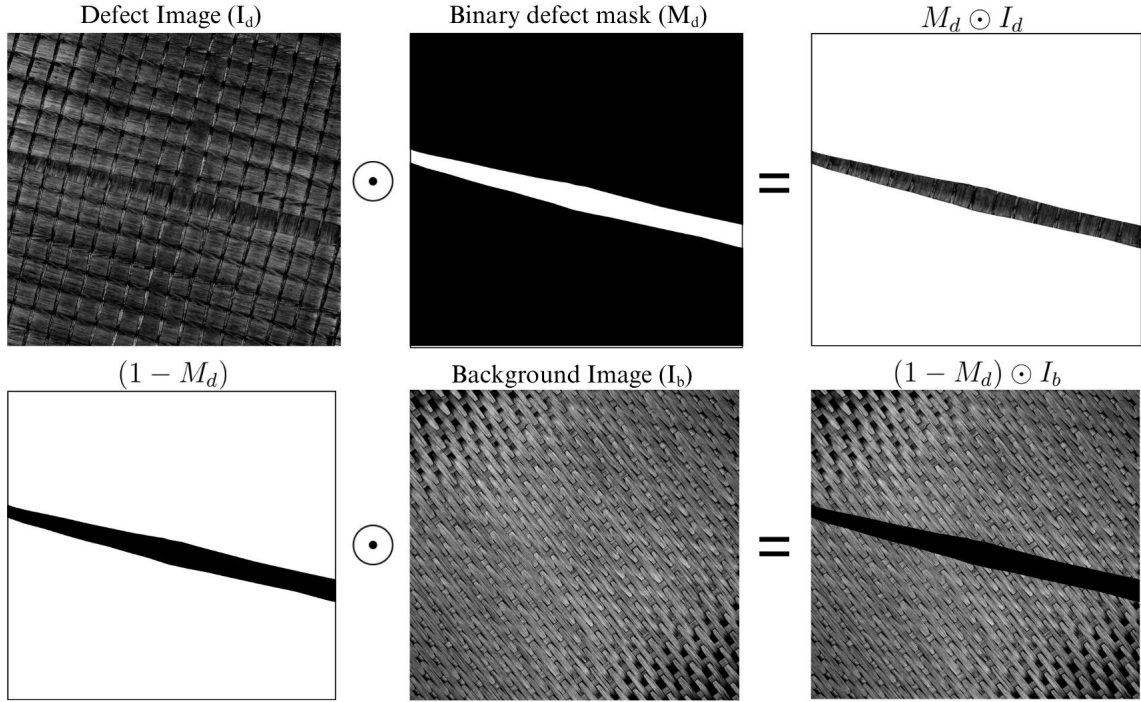
Leakage-safe augmentation strategy

A critical aspect of the CutMix implementation concerns the prevention of information leakage between training and validation subsets.

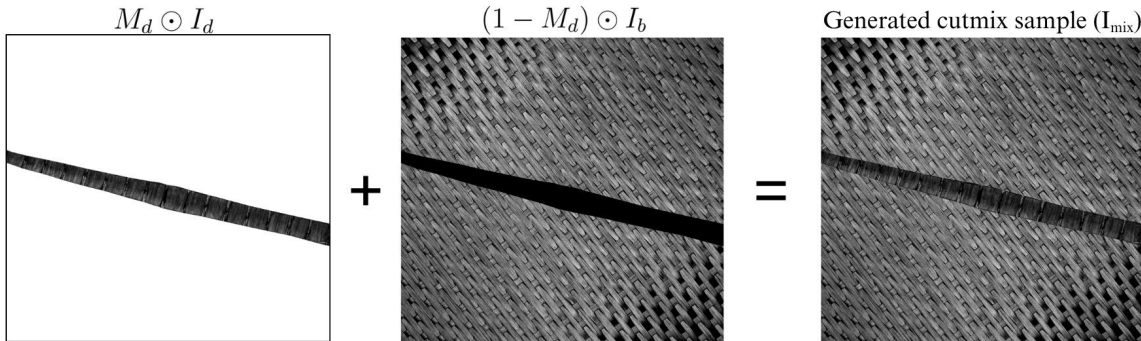
As discussed extensively in Chapter 4, grouped industrial datasets contain strongly correlated acquisitions derived from the same physical defect instances and material regions. Conse-

quently, augmentation strategies must be designed carefully, since synthetic sample generation can unintentionally violate the isolation between training and validation data even when the original dataset split itself is leakage-safe.

This issue is particularly relevant for CutMix-style augmentation. If defect donor samples or background textures originating from the validation split were reused during augmentation generation, the resulting synthetic training samples could indirectly expose the model to visual structures associated with the validation data. Under severe data scarcity conditions, this type



(a) Visual illustration of the masking operations involved in the generation of a CutMix sample. The donor defect image I_d is combined with the binary defect mask M_d to isolate the defect region, while the complementary mask $(1 - M_d)$ is applied to the clean background image I_b .



(b) Generation of the final CutMix sample I_{mix} through the combination of the masked defect region and the complementary background contribution.

Figure 5.5: Visual representation of the CutMix generation process adopted in this work.

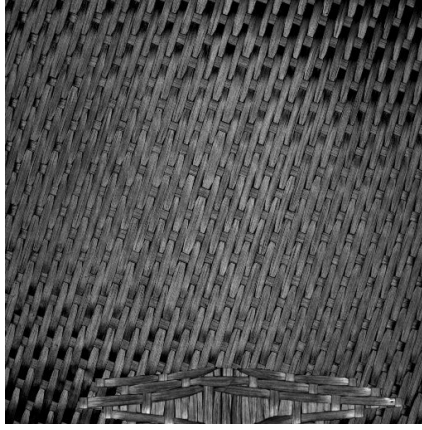


Figure 5.6: Example illustrating the effect of fibre-orientation mismatch during CutMix augmentation.

of indirect leakage could artificially inflate the measured generalization performance.

To preserve strict leakage-safe evaluation, CutMix generation was therefore constrained to operate exclusively within the training partition of each Monte Carlo split. Specifically:

- defect donor images were sampled exclusively from the training split;
- clean background images were selected only from materials not appearing in the corresponding validation split.

Consequently, no visual information originating from validation samples can be introduced into the augmented training set, either directly or indirectly. The resulting augmentation pipeline therefore preserves the same group-consistent leakage-safe assumptions adopted throughout the entire experimental methodology.

Challenges specific to carbon-fibre textures

Applying CutMix to carbon-fibre materials introduces additional challenges compared to standard natural-image augmentation settings.

Fibre orientation consistency. The visual appearance of defects strongly depends on the local orientation of the fibre bundles. When a defect extracted from one material orientation is pasted onto a background with a substantially different fibre direction, the resulting composite may appear physically inconsistent. Figure 5.6 illustrates an example of orientation mismatch during CutMix augmentation.

Missing stitch defects. A second limitation concerns the *missing stitch* defect class. In these samples, the annotated region corresponds to the absence of the stitching thread. Consequently, the transferred patch contains primarily the underlying textile texture rather than an



Figure 5.7: Example of CutMix augmentation for a missing stitch defect.

additional foreign structure. When such regions are pasted onto another material, the generated composite may therefore appear visually unusual. Figure 5.7 shows a representative example.

No explicit filtering strategy was implemented to remove physically implausible generated samples. In principle, certain inconsistencies could be detected and excluded during augmentation generation. For example, missing-stitch regions may transfer only the underlying textile texture rather than a distinct defective structure, while substantial orientation mismatches between donor and background textures may produce visually unrealistic composites. However, designing a reliable plausibility-selection strategy for industrial textile textures is non-trivial. Determining whether a generated sample is physically plausible would require additional constraints on fibre orientation consistency, local texture continuity, and material appearance compatibility. Moreover, under severe data scarcity conditions, aggressively discarding generated samples may unnecessarily reduce the effective augmentation diversity available during training. For this reason, the augmentation pipeline intentionally favours simplicity and diversity over strict physical realism.

In practice, strongly implausible composites often introduce highly visible texture discontinuities that are relatively easy for the network to identify. Consequently, although such samples may contribute less useful training information than physically realistic augmentations, their presence does not necessarily prevent the augmentation strategy from providing overall performance improvements.

Plausible generated sample. Despite these limitations, many generated samples remain visually plausible, particularly when fibre orientation and local texture are compatible between donor and background images. A representative example is shown in Figure 5.8.

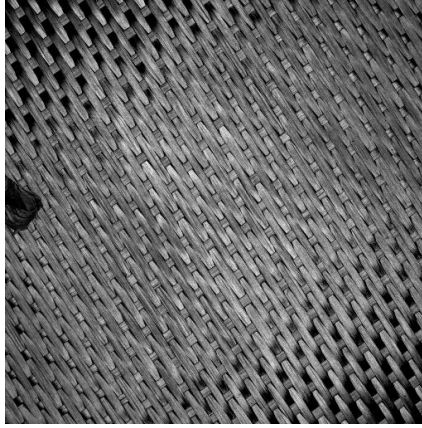


Figure 5.8: Example of physically plausible CutMix-generated composite.

5.7 Alternative Approaches Explored

In addition to the primary SegFormer-based semantic segmentation pipeline, several alternative approaches and exploratory strategies were investigated throughout the development of this work. The objective of these experiments was to assess whether different learning paradigms, transfer-learning strategies, or detection formulations could alleviate some of the limitations imposed by the scarcity of independent industrial defect samples. In particular, the following approaches were explored:

- PatchCore anomaly detection;
- YOLO-based instance segmentation;
- cross-dataset transfer learning using an external carbon-fibre inspection dataset.

Although these approaches were not retained as part of the final pipeline, they provide useful insights into the challenges associated with industrial carbon-fibre defect localization under limited-data conditions.

5.7.1 PatchCore anomaly detection

In addition to supervised semantic segmentation approaches, PatchCore [17] was investigated as a representative industrial anomaly-detection framework.

Unlike supervised segmentation methods, PatchCore operates using only clean (defect-free) training samples. The method constructs a memory bank of feature embeddings extracted from normal training images and identifies anomalies at inference time through nearest-neighbor distances in the feature space. More specifically, deep feature representations are extracted from a pretrained backbone network, such as ResNet, EfficientNet, or Vision Transformers. Spatial feature embeddings extracted from selected intermediate layers are partially

subsampled and stored in a memory bank representing the distribution of normal textures. During inference, feature embeddings extracted from a test image are compared against the stored memory-bank representations. Regions associated with large nearest-neighbor distances are interpreted as anomalous and contribute to the generation of an anomaly map.

Anomaly-detection approaches are particularly attractive in industrial inspection scenarios where collecting large numbers of pixel-level defect annotations is expensive and time-consuming. Training exclusively on clean samples also avoids the need to model the potentially large variability of defect appearances explicitly. For these reasons, anomaly detection represents a potentially appealing alternative under severe industrial data scarcity.

PatchCore is particularly effective in industrial inspection scenarios characterized by relatively homogeneous background structures and localized anomalous regions. In such settings, anomalies correspond to deviations from an otherwise consistent appearance distribution.

However, carbon-fibre materials present several characteristics that make this assumption less suitable. First, the entire image is composed of highly structured and repetitive textures rather than of a homogeneous background and a well-defined foreground object. Defects therefore appear as local perturbations of a complex periodic structure rather than as isolated foreign objects. Second, the background texture itself carries essential information for defect identification. Consequently, the boundary between normal texture variability and true anomalous structure becomes significantly less well-defined than in standard anomaly detection scenarios. In practice, the variability of normal textile appearance may itself become comparable to the visual perturbation introduced by subtle defect regions. A further challenge concerns sensitivity to fibre orientation. Since standard convolutional and transformer-based backbones are neither rotation invariant nor fully rotation equivariant, identical textures observed under different orientations may produce substantially different feature embeddings.

As a result, clean samples acquired under fibre orientations not well represented in the memory bank may themselves be **classified as anomalous**. Although rotation-equivariant architectures could partially mitigate this issue, reliable nearest-neighbor matching would still require the memory bank to contain clean samples spanning a large range of possible fibre orientations. Conversely, enforcing complete rotation invariance may suppress structural orientation information that is itself important for distinguishing defects from normal material organization. Additional limitations also arise from the dependence of the memory-bank representation on dataset curation, illumination conditions, acquisition setup, and material variability. In particular, materials or textures not sufficiently represented in the memory bank may be incorrectly identified as anomalous even in the absence of defects.

For these reasons, the assumptions underlying memory-bank anomaly detection appear only partially compatible with the highly structured and orientation-dependent nature of

carbon-fibre textures. PatchCore was therefore investigated primarily as an exploratory comparison with supervised segmentation approaches rather than as a direct replacement for the proposed representation-learning pipeline.

5.7.2 YOLO-based instance segmentation

In addition to semantic segmentation architectures based on the SegFormer framework, YOLO-based instance segmentation models were also considered as an alternative approach for industrial defect localization. YOLO models are widely adopted in industrial computer vision due to their computational efficiency and strong performance in object detection and instance segmentation tasks. Unlike semantic segmentation approaches, which directly predict a dense pixel-wise probability map over the entire image, YOLO-based segmentation models first detect object instances and subsequently predict an individual segmentation mask for each detected object.

YOLO segmentation architecture

The experiments in this work used the segmentation framework provided by the Ultralytics implementation. Two model variants were considered:

- YOLO26m segmentation model;
- YOLO26l segmentation model.

During training, YOLO jointly optimizes three tasks:

- bounding-box localization;
- object classification;
- instance segmentation.

Instead of directly predicting a full-resolution segmentation mask for each detected object, the segmentation branch predicts a set of object-specific coefficients used to combine a collection of shared prototype masks learned by the network. The final object mask is obtained through the weighted combination of these prototype representations. Consequently, the output of the YOLO segmentation architecture consists of:

- a bounding box identifying the predicted defect location;
- a confidence score associated with the detection;
- a dense segmentation mask describing the defect shape.

This differs fundamentally from the semantic segmentation formulation adopted by SegFormer, where the network predicts a dense probability map over all image pixels without requiring explicit object detections.

Confidence-threshold considerations

YOLO-based models rely on confidence thresholds in order to filter low-confidence detections during inference. In standard detection pipelines, predictions whose confidence falls below a predefined threshold are discarded. In practice, this reduces false positives and stabilizes the final predictions. However, this mechanism also introduces an important conceptual difference with respect to semantic segmentation architectures. In semantic segmentation, every pixel receives a prediction independently of object-level confidence. In contrast, YOLO first determines whether a defect instance is present and only then generates the associated segmentation mask. For this reason, removing the confidence threshold entirely may lead to severe over-segmentation, since low-confidence detections are no longer suppressed. This distinction becomes particularly relevant in industrial defect inspection scenarios involving subtle texture anomalies and highly imbalanced foreground/background distributions.

The effectiveness of YOLO-based instance segmentation under the considered industrial inspection setting is evaluated experimentally and compared with semantic segmentation approaches in Chapter 6.

5.7.3 Cross-dataset transfer-learning exploration

Given the limited number of independent defect instances available in the proposed dataset, an additional transfer-learning experiment was conducted to evaluate whether intermediate fine-tuning on an external carbon-fibre inspection dataset could improve segmentation performance and feature robustness. The external dataset was provided by the Bristol Robotics Laboratory and was introduced previously in the related-work discussion of Chapter 2. The corresponding study investigated SegFormer-based segmentation of carbon-fibre surface defects under controlled illumination conditions and discussed challenges closely related to those addressed in this thesis, including limited dataset size, domain-specific acquisition settings, and constrained generalization. Since both studies address semantic segmentation of carbon-fibre surface defects using transformer-based architectures, intermediate fine-tuning on the Bristol dataset was considered a potentially promising strategy for improving feature robustness before specialization on the proposed dataset.

Motivation. The motivation behind this experiment was based on the hypothesis that an intermediate adaptation stage on a related industrial inspection dataset could help the encoder learn more suitable texture representations before fine-tuning on the target domain. More specifically, the objective was to evaluate whether pre-adaptation to carbon-fibre inspection imagery could help bridge the substantial domain gap between natural-image pretraining and the target task, reducing the amount of domain adaptation required from ADE20K-pretrained

weights alone. Under severe data-scarcity conditions, a significant portion of the available training data may otherwise be effectively used to adapt low-level representations to the industrial domain before learning task-specific defect characteristics. The experiment therefore aimed to determine whether a two-stage fine-tuning procedure could improve generalization by exposing the model to a larger amount of industrial texture data before the final adaptation stage on the proposed dataset.

Acquisition setup and selected subset. The Bristol dataset consists primarily of grayscale industrial inspection images acquired under several controlled illumination configurations. The original study evaluated multiple acquisition setups, including directional illumination and dome-light illumination conditions. For the present experiment, the dome-light subset was selected, since it produced the best segmentation performance in the original Bristol experiments and provided visually more stable surface representations. The Bristol dataset differs significantly from the proposed dataset not only in terms of acquisition setup, but also in terms of the image representation itself. Whereas the proposed pipeline relies on a custom three-channel representation composed of azimuthal, fibular, and low-pass-filtered diffuse channels derived from the photometric acquisition system, the Bristol dataset consists primarily of single-channel grayscale intensity images. In order to preserve compatibility with the three-channel input projection layer of the SegFormer encoder during the intermediate fine-tuning stage, the grayscale images of the Bristol dataset were replicated across three channels before training. A similar strategy was also adopted by the authors of the Bristol dataset in their original work, where the grayscale textile images were duplicated across the three input channels in order to maintain compatibility with architectures pretrained on RGB natural-image datasets.

Consequently, the three input channels remain perfectly correlated during the intermediate adaptation stage and do not provide the complementary multi-modal information available in the proposed sensor-derived representation pipeline. Although this preserves architectural compatibility, the resulting input representation remains fundamentally different from the sensor-derived representations adopted in the proposed pipeline.

Annotation strategy and task formulation. An additional difference concerns the annotation strategy adopted in the Bristol study. The original Bristol experiments considered two separate segmentation scenarios:

- segmentation of general surface defects;
- segmentation of foreign objects.

These two tasks were treated independently in the original work through separate annotation configurations. For the present experiment, only the annotations corresponding to sur-

face defects were retained, while foreign-object annotations were excluded in order to maintain greater consistency with the target defect-segmentation task considered in this thesis. Nevertheless, the defect taxonomy and visual morphology of the annotated structures remain substantially different from those present in the proposed dataset.

Consequently, the intermediate fine-tuning stage does not correspond to direct task continuation, but rather to partial transfer between related industrial texture-analysis domains characterized by different defect morphologies and acquisition statistics.

Transfer-learning procedure. The adopted transfer-learning strategy consisted of two consecutive fine-tuning stages:

1. initial fine-tuning on the Bristol Robotics Laboratory dataset;
2. subsequent fine-tuning on the proposed dataset using the same training configuration adopted in the main representation-analysis experiments.

The second fine-tuning stage therefore reused the same representation design, normalization pipeline, and optimization configuration adopted in the principal SegFormer experiments of this thesis. The underlying hypothesis was that intermediate adaptation to a related industrial texture-inspection task could provide more robust feature representations and improve generalization under severe data-scarcity conditions.

Whether this intermediate industrial adaptation stage effectively improves downstream segmentation performance remains an empirical question investigated experimentally in Chapter 6.

5.8 Chapter summary

This chapter presented the methodological framework adopted for the experimental analysis of industrial carbon-fibre defect segmentation.

Particular attention was devoted to the challenges introduced by severe data scarcity, grouped correlated acquisitions, and the non-standard sensor representations provided by the industrial inspection system. To address these issues, the chapter introduced a leakage-safe evaluation methodology together with a representation-design pipeline specifically tailored to carbon-fibre texture analysis. Several complementary representation strategies were investigated, including sensor-channel selection, azimuthal orientation mappings, frequency-domain filtering, and normalization procedures designed to improve compatibility with RGB-pretrained transformer backbones. The chapter also discussed augmentation strategies, transfer-learning approaches, optimization configurations, and alternative anomaly detection frameworks explored throughout the experimental analysis. The resulting methodology establishes the foundation for the comparative experiments presented in Chapter 6, where the influence of the proposed design choices is evaluated quantitatively and qualitatively under leakage-safe industrial evaluation conditions.

6 Experimental Results

This chapter presents the experimental analysis conducted using the methodological framework introduced in the previous chapter under the leakage-safe evaluation protocol described in Chapter 4. The objective of the experiments is to analyse how different methodological choices influence segmentation performance and generalization behaviour under realistic industrial constraints characterized by limited data availability, strong texture variability, correlated multi-acquisition observations, and non-standard sensor-derived representations. The chapter is organized as a sequence of controlled experimental studies investigating:

- sensor-representation design;
- azimuthal normalization strategies;
- frequency-domain preprocessing;
- backbone-capacity scaling;
- optimization and regularization strategies;
- transfer-learning configurations;
- and CutMix-based data augmentation.

Most configurations are evaluated using ten Monte Carlo seeds in order to reduce the variance induced by the limited number of independent defect instances available in the dataset. Some intermediate hyperparameter explorations employ three seeds sampled from the same Monte Carlo pool in order to accelerate the experimental cycle while preserving comparability within the same subsection. It is important to note that the quantitative results reported across different subsections are not always directly comparable, since several later experiments adopt improved representations, optimization procedures, or regularization strategies identified during earlier stages of the analysis. Consequently, the primary objective of this chapter is to analyse the relative influence of individual methodological choices under controlled evaluation conditions rather than to establish a single global ranking across all configurations.

6.1 Input Representation Analysis

The first experimental study investigates how different sensor-representation strategies influence segmentation performance under leakage-safe industrial evaluation conditions. The experiments reported in this section modify a single component of the representation pipeline at a time while preserving the remaining training configuration unchanged. This controlled evaluation strategy allows the influence of each methodological choice to be analysed in isolation.

As discussed in Chapter 5, the Fscan inspection system provides multiple feature channels describing complementary physical properties of the carbon-fibre surface, including orientation-related and reflectance-related information. Since SegFormer is pretrained on RGB natural images, determining how these non-standard sensor channels should be integrated into the input representation becomes an important methodological question.

More specifically, the following experimental analyses investigate:

- the influence of input dimensionality and channel selection;
- the effect of normalization strategies;
- the usefulness of frequency-based representations;
- and the influence of alternative azimuthal mapping strategies.

6.1.1 Input dimensionality

As discussed in Section 5.1.2, an important representation-design question concerns the number of sensor channels provided as input to the segmentation network.

Since SegFormer is originally designed for three-channel RGB images, two principal strategies were considered:

- modifying the input projection layer to accept all five available channels;
- selecting a subset of three informative channels while preserving compatibility with the pretrained input structure.

The experiments reported in this subsection isolate the influence of input dimensionality while preserving the same normalization strategy and training configuration across all runs. Table 6.1 summarizes the obtained results. The results indicate that increasing the number of input channels from three to five does not produce a measurable improvement in segmentation performance under the considered experimental setting. This observation is particularly important because the five-channel representation introduces a higher-dimensional input space while also requiring modification of the pretrained input projection layer. Despite the increased representational capacity, the obtained performance remains essentially unchanged. Under severe

Input representation	Hard IoU
A/F/D (three channels)	0.5008 ± 0.0791
A/F/D/P/S (five channels)	0.5007 ± 0.0738

Table 6.1: Influence of input dimensionality on segmentation performance. Results are reported as mean $\pm$ sample standard deviation across 10 Monte Carlo seeds.

data scarcity, the available dataset may simply be too small for the model to effectively adapt to the higher-dimensional representation. In contrast, the three-channel configuration preserves direct compatibility with RGB-pretrained architectures while achieving equivalent segmentation performance with lower representational complexity. This does not imply that five-channel representations are intrinsically inferior. With substantially larger industrial datasets, the additional sensor channels could potentially provide useful complementary information. However, within the constrained setting considered in this work, the simpler three-channel representation provided a more practical and computationally efficient solution.

The relatively large standard deviations observed in both configurations motivate a more detailed discussion of variability under leakage-safe evaluation conditions.

6.1.2 Variability under leakage-safe evaluation

As discussed in Chapter 5, all experiments in this work are evaluated using repeated Monte Carlo sampling in order to reduce the sensitivity of the results to a specific train-validation partition. Under severe industrial data scarcity conditions, the segmentation performance may vary substantially depending on which defect instances are available during training. This effect becomes particularly significant under grouped leakage-safe splitting, where the validation set contains entirely unseen physical defect instances rather than correlated acquisitions of defects already observed during training.

The proposed dataset contains a limited number of independent defect groups together with substantial variability in defect morphology, material appearance, fibre orientation, and recycled-material texture characteristics. Consequently, different Monte Carlo seeds may generate training partitions with substantially different levels of difficulty. This variability is reflected directly in the obtained experimental results.

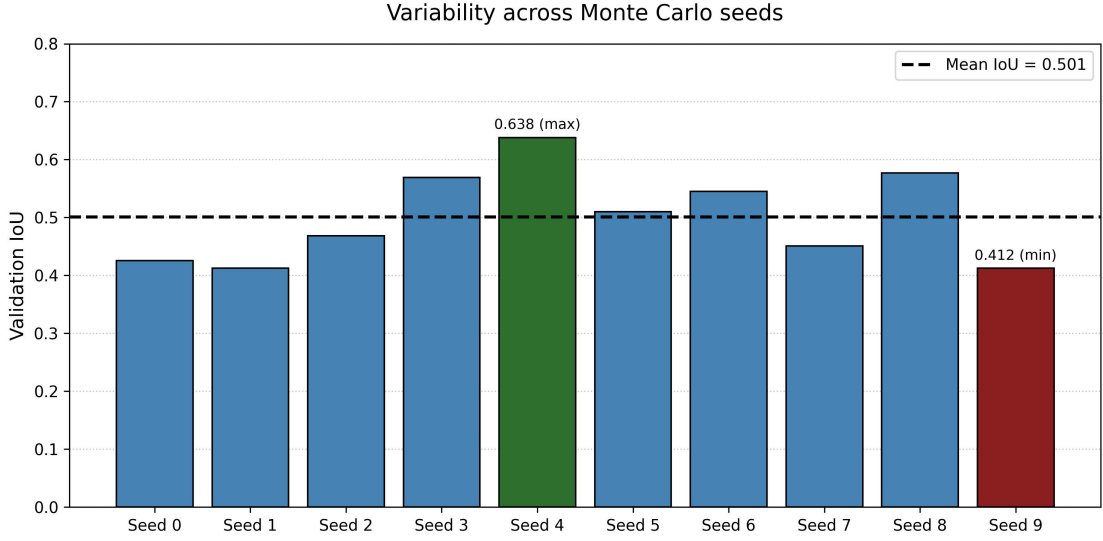


Figure 6.1: Validation IoU obtained across the ten Monte Carlo seeds for the initial three-channel input configuration under leakage-safe instance-grouped evaluation. Although the architecture and training configuration remain identical, substantial variability emerges depending on the specific train–validation partition generated by each seed. The best-performing seed achieved an IoU of 0.638, whereas the worst-performing seed achieved only 0.412. This behaviour highlights the strong sensitivity of industrial defect segmentation to data partitioning under severe data scarcity and motivates the use of repeated Monte Carlo evaluation instead of single-split performance estimation.

For example, in the initial input-dimensionality experiment, the same three-channel configuration produced validation IoU values ranging from:

$$0.412 \leq \text{IoU} \leq 0.638$$

across different Monte Carlo seeds.

Similarly, the five-channel configuration produced results ranging from:

$$0.404 \leq \text{IoU} \leq 0.606.$$

These observations highlight that evaluating a single random train–validation partition may produce either overly optimistic or overly pessimistic conclusions regarding the true generalization capability of the model. For this reason, the experimental analysis reported throughout this chapter relies primarily on the mean and standard deviation across multiple Monte Carlo runs rather than on single-split performance estimates. Figure 6.1 illustrates the variability observed across the ten Monte Carlo seeds for the initial three-channel configuration.

These observations indicate that performance variability is itself an important characteristic of the considered industrial evaluation setting rather than merely a source of experimen-

tal noise. Consequently, reliable assessment of segmentation performance requires repeated leakage-safe evaluation together with statistical aggregation across multiple Monte Carlo runs.

6.1.3 Effect of normalization strategies

As discussed in Section 5.2, additional normalization stages may improve compatibility between sensor-derived representations and the RGB-pretrained SegFormer encoder. Three normalization configurations were therefore compared:

- no additional normalization;
- per-sample min–max normalization;
- per-sample min–max normalization followed by ImageNet normalization.

The experiments were conducted using the same A/F/D channel representation and identical training settings across all runs. Table 6.2 summarizes the obtained results.

Normalization strategy	Hard IoU	Precision	Recall
No normalization	0.501 ± 0.079	0.710 ± 0.06	0.630 ± 0.09
Min–max normalization	0.521 ± 0.079	0.708 ± 0.08	0.669 ± 0.10
Min–max + ImageNet normalization	0.535 ± 0.077	0.705 ± 0.07	0.690 ± 0.09

Table 6.2: Influence of normalization strategies on segmentation performance. Results are reported as mean $\pm$ sample standard deviation across 10 Monte Carlo seeds.

The results indicate that normalization consistently improves the segmentation performance. In particular, the combination of per-sample min–max normalization and ImageNet normalization produces the best overall results. Interestingly, the improvement associated with normalization is reflected primarily in recall rather than precision. This behaviour suggests that normalization improves the ability of the model to identify defect regions more completely, while preserving a comparable level of false-positive control.

From an industrial inspection perspective, this behaviour is particularly desirable since missed defects are generally more costly than moderate over-segmentation. This behaviour is consistent with the fact that the SegFormer encoder is pretrained on ImageNet and therefore expects inputs following activation statistics compatible with natural-image training. Even though the semantic meaning of the sensor-derived channels differs from RGB images, aligning the numerical distribution of the inputs with the pretrained backbone improves optimization stability and feature reuse.

The publicly available SegFormer checkpoints used in this work are fine-tuned on ADE20K semantic segmentation. However, the encoder weights ultimately originate from ImageNet

pretraining, and the standard preprocessing pipeline associated with these checkpoints still relies on ImageNet normalization statistics. Moreover, the objective of this normalization stage is not to perfectly reproduce the statistics of the ADE20K dataset itself, but rather to maintain compatibility with the activation distributions expected by the pretrained encoder during transfer learning. The usefulness of this normalization strategy was therefore treated as an empirical question and evaluated experimentally in the present subsection.

6.1.4 Frequency-based representation study

Frequency-domain preprocessing may help suppress fine-grained periodic textile patterns while preserving larger structural organization associated with defect regions. The baseline configuration corresponds to the A/F/D representation combined with min–max and ImageNet normalization. Low-pass and high-pass filtered variants of the diffuse channel were compared against this baseline representation. The results indicate that low-pass filtering of the diffuse

Input representation	Hard IoU
A/F/D	0.535 ± 0.077
A/F/LPF(diffuse)	0.560 ± 0.064
A/F/HPF(diffuse)	0.522 ± 0.062

Table 6.3: Influence of frequency-based representations on segmentation performance. Results are reported as mean $\pm$ sample standard deviation across 10 Monte Carlo seeds.

channel provides the best overall performance among the evaluated representations. This observation suggests that preserving large-scale structural organization is beneficial for defect localization in carbon-fibre textures. By attenuating fine-grained periodic texture components, low-pass filtering appears to emphasize broader structural discontinuities associated with defect regions. In contrast, representations based primarily on high-frequency information do not provide comparable improvements. Although high-pass filtering highlights local texture irregularities, isolating these components alone appears insufficient for robust segmentation under severe data scarcity. Overall, the experiments suggest that defect localization benefits more from preserving broader structural organization than from emphasizing only local high-frequency variations.

6.1.5 Azimuthal channel mapping

The influence of the azimuthal channel mapping strategy introduced in Section 5.2.1 was evaluated experimentally. The numerical values reported in this subsection should not be compared directly with those of the previous representation experiments, since the present analysis was

conducted using only three Monte Carlo seeds. The results are therefore intended primarily for relative comparison between the two azimuthal mapping strategies.

Two alternatives were compared:

- a directional mapping preserving the full angular direction;
- an orientation-only mapping collapsing opposite directions through a modulo- π transformation.

The objective of this experiment was to evaluate whether preserving directionality provides useful information for defect localization or whether the physically undirected nature of carbon-fibre tows makes directional information unnecessary.

Azimuthal mapping	Hard IoU
Directional mapping	0.602 ± 0.037
Orientation-only mapping	0.574 ± 0.067

Table 6.4: Influence of the azimuthal channel mapping strategy. Results are reported as mean $\pm$ sample standard deviation across 3 Monte Carlo seeds. *The numerical values are intended only for relative comparison within this experiment.*

The results indicate that preserving directional information produces slightly better segmentation performance. Although opposite fibre directions are physically equivalent from a material-orientation perspective, preserving directional encoding may still provide additional spatial continuity cues that help the network model local structural variations associated with defect regions. These observations suggest that the artificial discontinuities introduced by directional encoding do not outweigh the additional information preserved by the representation.

6.2 Backbone Capacity Comparison

All representation experiments discussed in the previous sections were performed using the SegFormer-B2 backbone. This intermediate-capacity configuration was adopted initially in order to isolate the influence of representation and normalization strategies while keeping the architectural complexity fixed. After identifying the most effective representation pipeline, the influence of backbone capacity was investigated explicitly through a comparative study of three SegFormer variants with progressively increasing encoder size: SegFormer-B0, SegFormer-B2, and SegFormer-B5. The objective of this experiment is to evaluate how model capacity influences segmentation performance and generalization behaviour under severe data scarcity conditions.

Quantitative comparison

The results reported in Table 6.5 show a clear improvement when moving from the lightweight B0 backbone to the medium-capacity B2 configuration.

Backbone	Hard IoU	Precision	Recall
SegFormer-B0	0.467 ± 0.078	0.652 ± 0.084	0.626 ± 0.090
SegFormer-B2	0.560 ± 0.064	0.743 ± 0.061	0.694 ± 0.072
SegFormer-B5	0.557 ± 0.082	0.715 ± 0.080	0.715 ± 0.077

Table 6.5: Influence of backbone capacity on segmentation performance. Results are reported as mean $\pm$ sample standard deviation across 10 Monte Carlo seeds.

Both segmentation overlap and detection-oriented metrics improve substantially, suggesting that the B0 architecture is likely under-parameterized for the complexity of the considered industrial segmentation task. Increasing the capacity further to the B5 variant does not produce additional improvements in Hard IoU. Interestingly, however, the larger backbone exhibits a slightly different precision–recall trade-off. While the B2 configuration achieves the highest precision, the B5 backbone produces the highest recall. This behaviour suggests that the larger model may become more sensitive to subtle defect-like structures, increasing its ability to identify defect regions while also introducing additional false positives. Overall, the results indicate that the B2 backbone provides the most favourable balance between representational capacity, optimization stability, and generalization under severe data scarcity conditions.

Qualitative comparison

Figure 6.2 shows representative qualitative segmentation results obtained with the three backbone variants. The B2 and B5 models generally produce more coherent and complete segmentations than the lightweight B0 backbone. However, the qualitative differences between B2 and B5 remain relatively limited, which is consistent with the quantitative results and suggests diminishing returns from increasing backbone capacity under severe data scarcity.

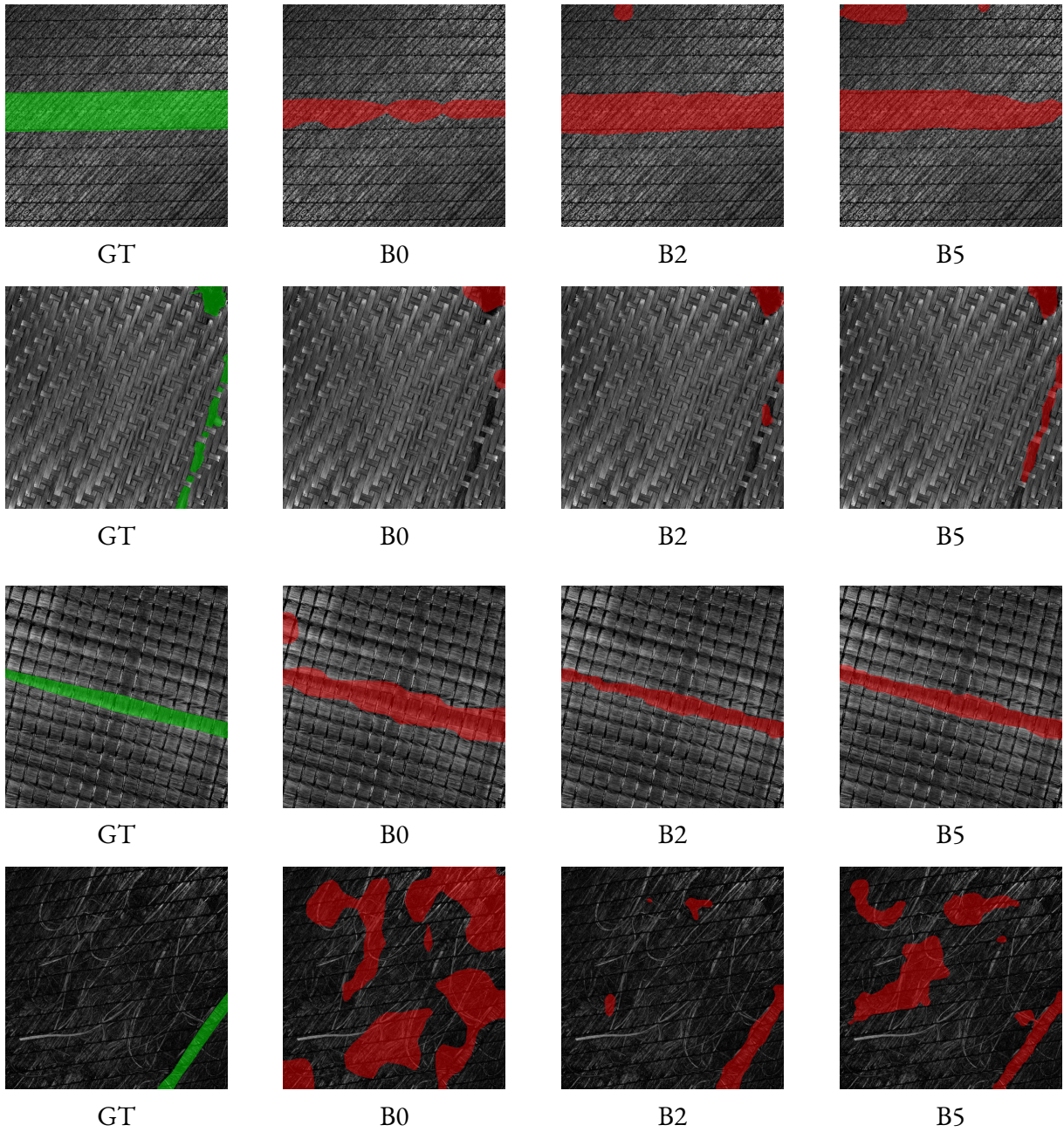


Figure 6.2: Qualitative comparison of segmentation outputs for different SegFormer backbone capacities. The first column shows the input image with the ground-truth annotation overlay (green). The following columns show the predictions of the B0, B2, and B5 backbones respectively (red).

6.3 CutMix augmentation study

After identifying the most effective representation and backbone configuration, additional experiments were conducted to investigate whether CutMix augmentation could improve segmentation performance under severe data scarcity conditions.

The experiments compare the same baseline configuration trained with and without CutMix augmentation while preserving identical grouped train–validation partitions and Monte Carlo seeds. As discussed in Section 5.6, the augmentation pipeline was explicitly designed to preserve leakage-safe evaluation. All synthetic samples were generated exclusively from training-set data within each Monte Carlo split, while the validation subsets remained completely unchanged. For the augmentation experiments, each defective sample was allowed to generate at most 20 CutMix composites using randomly selected defect-free background textures whose material type did not appear in the corresponding validation subset. This procedure produced a training set containing slightly more than 6000 generated augmented samples.

6.3.1 Quantitative results

The results in Table 6.6 indicate that CutMix augmentation produces a consistent improvement in segmentation performance.

Configuration	Hard IoU	Precision	Recall
A/F/LPF(diffuse)	0.56 ± 0.064	0.74 ± 0.06	0.69 ± 0.07
A/F/LPF(diffuse) + CutMix	0.63 ± 0.079	0.81 ± 0.06	0.75 ± 0.08

Table 6.6: Comparison between the baseline configuration and the same configuration trained with CutMix augmentation.

Compared with the baseline configuration, the CutMix model improves Hard IoU from 0.56 to 0.63, while also increasing both precision and recall. This suggests that the augmentation does not merely increase the number of predicted defect pixels, but improves the overall balance between defect localization and false-positive control. Although the augmentation substantially increases the nominal size of the training dataset, the improvement remains moderate relative to the number of generated samples. As discussed in Section 5.6, some composites may be only partially physically plausible because of fibre-orientation mismatches or texture inconsistencies between donor defects and background materials. Therefore, the effective value of the generated samples depends not only on their quantity, but also on their structural realism. Nevertheless, despite these limitations, the augmentation strategy still produces measurable improvements under severe data scarcity conditions.

6.3.2 Qualitative comparison

To complement this analysis, Fig. 6.3 shows representative qualitative comparisons between predictions obtained with and without CutMix augmentation.

For each example, the first column shows the ground-truth annotation, whereas the remaining columns show the predictions produced by the corresponding models.

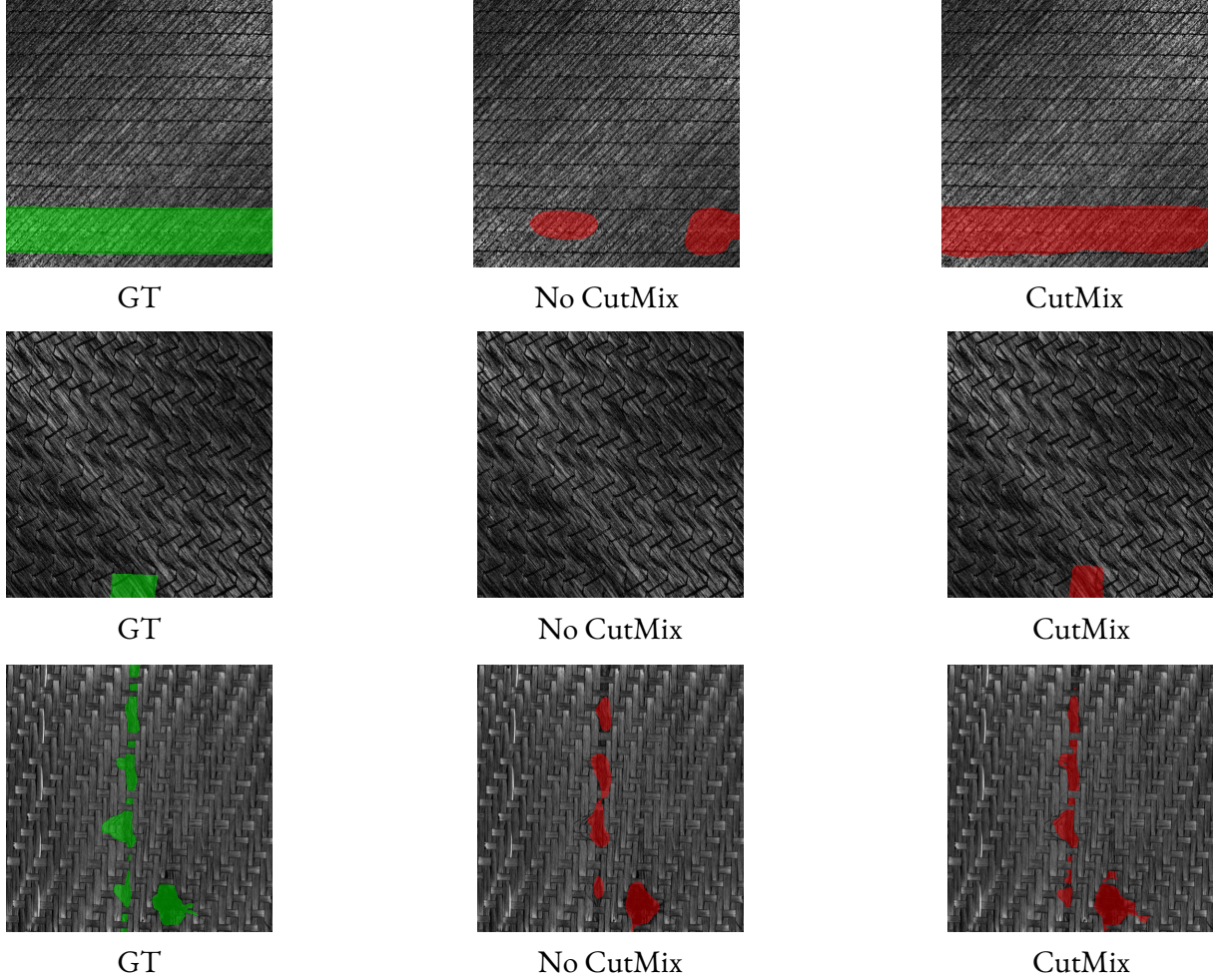


Figure 6.3: Qualitative comparison between models trained without and with CutMix augmentation.

6.3.3 Interpretation of the CutMix improvements

The CutMix experiments suggest that synthetic augmentation can partially compensate for the limited number of independent defect instances available during training. Both the quantitative and qualitative analyses indicate that exposing the network to a wider variety of defect-background combinations improves segmentation robustness and reduces failure cases associated with incomplete detections. At the same time, the observed improvement remains moderate relative to the large number of generated samples, indicating that the usefulness of synthetic augmentation depends not only on the quantity of generated samples but also on their physical consistency and structural realism. Nevertheless, despite these limitations, the augmentation

strategy still produces measurable gains, further supporting the conclusion that data scarcity represents a major limiting factor for the industrial segmentation task considered in this thesis.

6.3.4 Scope of the supervised segmentation analysis

The experiments presented in the previous sections focused on the main supervised semantic segmentation framework investigated within the scope of this thesis, including representation design, normalization strategies, backbone selection, and leakage-safe augmentation. Taken together, these experiments demonstrated that careful methodological design can improve segmentation robustness under realistic industrial constraints, while also highlighting the persistent limitations imposed by severe data scarcity and material variability. The following section presents several additional exploratory approaches that were investigated in parallel to the main supervised segmentation pipeline.

6.4 Alternative approaches and exploratory experiments

In addition to the main supervised SegFormer-based segmentation framework analysed throughout the previous sections, several alternative approaches were also explored in order to assess different strategies for industrial defect localization and anomaly detection. These experiments were not intended to replace the primary segmentation pipeline, but rather to provide additional insight into different paradigms potentially applicable to industrial carbon-fibre inspection under severe data scarcity conditions. The following sections therefore present a concise overview of these exploratory investigations together with the main observations that emerged from them.

6.4.1 Comparison with YOLO segmentation models

Additional experiments were conducted using YOLO-based instance segmentation models in order to compare the semantic segmentation approach adopted by SegFormer with a detection-oriented formulation. Both YOLO variants were trained using the same Monte Carlo splits adopted for the SegFormer experiments, ensuring that the comparison was performed under identical data conditions.

Confidence-threshold behaviour

The default inference configuration adopted a confidence threshold of 0.25, corresponding to the standard Ultralytics setting. An additional experiment was performed using a confidence threshold of 0.0 in order to make the YOLO predictions more comparable to the dense

semantic predictions produced by SegFormer. However, removing confidence filtering produced a large number of spurious detections and severe over-segmentation of the textile background. In practice, the model tended to classify large portions of the carbon-fibre texture as defects. These observations highlight the strong dependence of YOLO-based architectures on confidence-based detection filtering.

Quantitative comparison

The quantitative results highlight substantial differences between the two segmentation strategies. YOLO achieved higher precision, indicating that predicted defect regions were often spatially accurate once detected. However, the substantially lower recall indicates that many defects were not detected at all. In contrast, SegFormer achieved a substantially better balance between precision and recall, resulting in significantly higher overall segmentation performance.

These observations suggest that dense semantic segmentation is more suited to the considered industrial inspection scenario, where missed defects are generally more critical than limited over-segmentation.

Model	Hard IoU	Precision	Recall
SegFormer	0.560 ± 0.064	0.74 ± 0.06	0.69 ± 0.07
YOLO26-M	0.280 ± 0.077	0.81 ± 0.07	0.30 ± 0.09

Table 6.7: Comparison between SegFormer and YOLO segmentation models evaluated on the same set of 10 Monte Carlo seeds.

Qualitative comparison

To better understand the differences highlighted by the quantitative results, Fig. 6.4 presents representative qualitative comparisons between YOLO predictions and the corresponding SegFormer outputs. For each example, the first column shows the ground-truth annotation, whereas the remaining columns show the YOLO bounding-box prediction, the YOLO segmentation mask, and the corresponding SegFormer prediction respectively.

The YOLO predictions generally exhibit sharper and more compact boundaries due to the instance-based formulation of the model. However, YOLO also produced substantially more missed detections, particularly for small or low-contrast defects. In contrast, the semantic segmentation strategy adopted by SegFormer tended to produce more spatially extensive predictions while achieving higher defect recall.

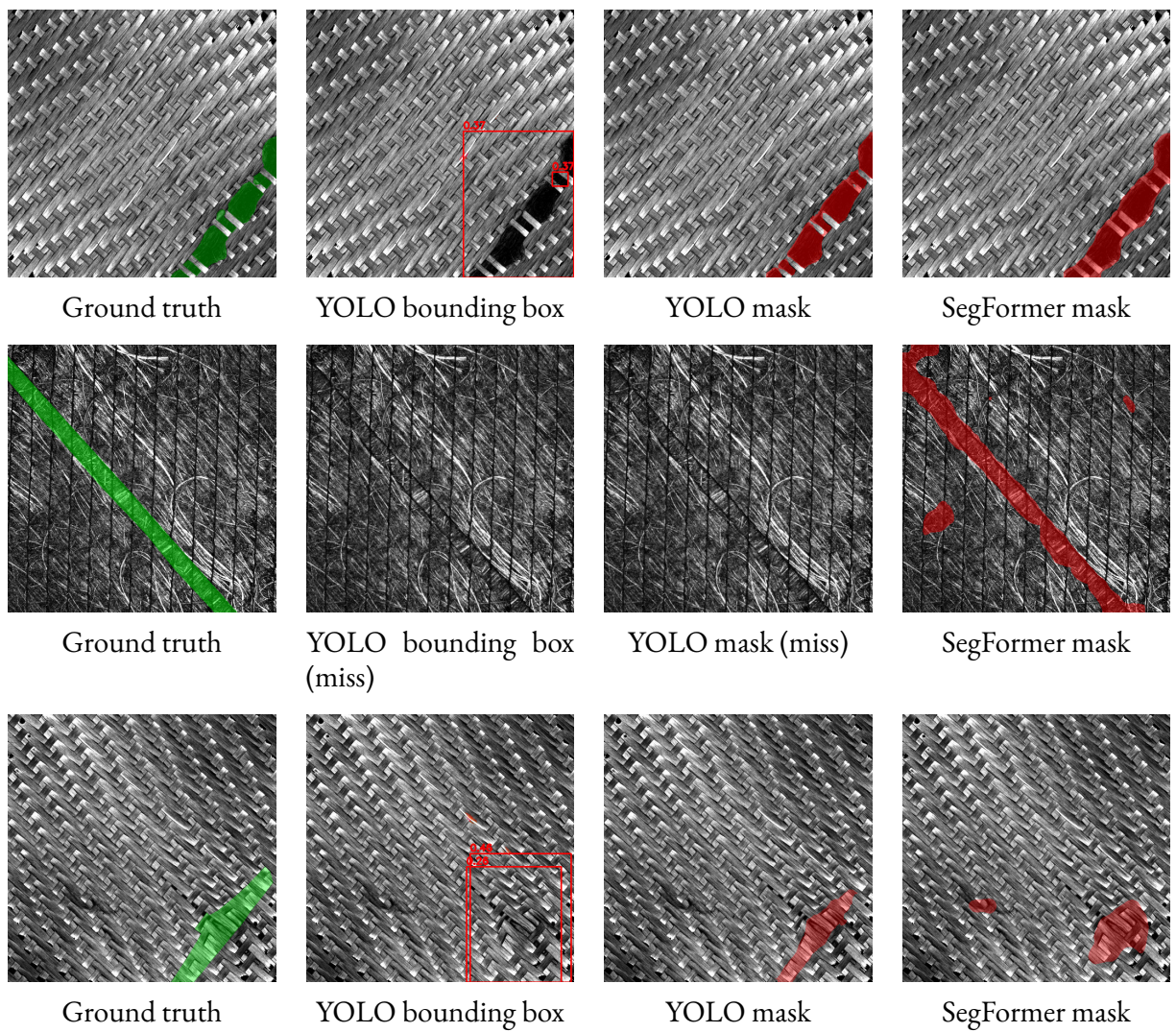


Figure 6.4: Qualitative comparison between YOLO instance segmentation and SegFormer semantic segmentation predictions.

6.4.2 PatchCore anomaly detection

PatchCore was evaluated qualitatively as an alternative anomaly detection framework for carbon-fibre defect localization. Unlike the supervised segmentation approaches adopted throughout the main experiments of this thesis, PatchCore relies exclusively on clean training samples and detects anomalies through nearest-neighbor distances in a learned feature space. In practice, the obtained results highlighted several limitations when applying anomaly-detection approaches to highly structured carbon-fibre textures. Figure 6.5 shows a representative anomaly map obtained on a gap defect.



Figure 6.5: PatchCore anomaly map on a gap defect.

Additional examples for fuzzball and missing-stitch defects are shown in Figures 6.6 and 6.7.



Figure 6.6: PatchCore anomaly map on a fuzzball defect.

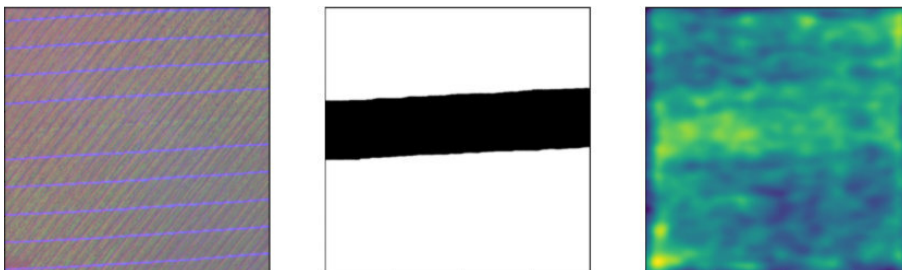


Figure 6.7: PatchCore anomaly map on a missing stitch defect.

Each qualitative example shown in the following figures consists of three panels. The left panel shows the input sample used by PatchCore, where the azimuthal, fibular, and diffuse

channels are combined into an RGB visualization for display purposes. Unlike the qualitative examples presented previously for SegFormer, where only the fibular channel was shown for simplicity, the available PatchCore visualizations were generated directly using this three-channel representation. The central panel shows the corresponding ground-truth annotation from the original industrial dataset. In the available annotations, the mask convention is inverted with respect to the standard semantic-segmentation representation, with defect regions encoded as 0 and background pixels encoded as 255. The right panel shows the anomaly map produced by PatchCore.

Quantitative evaluation was not performed for these exploratory PatchCore experiments. Since PatchCore produces continuous anomaly maps rather than explicit binary segmentations, obtaining segmentation metrics would require the introduction of an additional threshold-selection procedure for converting the anomaly responses into binary masks. Moreover, the qualitative analysis already suggested limited suitability of the approach for the considered industrial defect-segmentation setting, particularly in the presence of strong texture variability and subtle fibre-orientation changes.

The qualitative results suggest that PatchCore is capable of highlighting certain defect regions, particularly when the local texture disruption produces a strong deviation from the normal material structure. However, the anomaly maps also exhibit strong sensitivity to the underlying fibre organization and to local orientation changes. In several cases, portions of normal texture were assigned high anomaly scores despite not containing actual defects. These observations are consistent with the theoretical limitations discussed in Chapter 5. Since the carbon-fibre texture itself contains highly structured directional information, local feature variations associated with normal material organization may produce distances in feature space comparable to those generated by actual defects.

More generally, the results suggest that purely memory-based anomaly-detection approaches may struggle to distinguish between normal texture variability and genuine structural anomalies in highly repetitive industrial materials such as carbon-fibre surfaces. For this reason, supervised semantic segmentation approaches based on task-specific representation learning appeared more suitable for the defect-localization problem considered in this work.

6.4.3 Cross-dataset transfer-learning experiment

This experiment evaluated whether intermediate fine-tuning on the external Bristol Robotics Laboratory dataset could improve segmentation performance on the proposed dataset. The experiment adopted the transfer-learning procedure introduced in Section 5.7.3, consisting of an initial fine-tuning stage on the Bristol dataset followed by subsequent fine-tuning on the proposed industrial dataset. The same representation design and training configuration

adopted in the main representation-analysis experiments were preserved in order to isolate the effect of the intermediate adaptation stage.

Configuration	Hard IoU
Direct fine-tuning on proposed dataset	0.560 ± 0.064
Intermediate Bristol fine-tuning + proposed dataset	0.539 ± 0.076

Table 6.8: Influence of intermediate fine-tuning on the Bristol Robotics Laboratory dataset before adaptation to the proposed dataset. Results are reported as mean $\pm$ sample standard deviation across 10 Monte Carlo seeds.

Contrary to the initial expectations, the introduction of the intermediate fine-tuning stage did not improve segmentation performance.

The obtained results suggest that transferability between the two industrial datasets is more limited than initially anticipated, despite the apparent similarity of the application domain. One possible explanation concerns the substantial difference between the input representations used in the two datasets.

The Bristol dataset consists primarily of grayscale intensity images, whereas the proposed pipeline relies on a custom three-channel representation composed of azimuthal, fibular, and low-pass-filtered diffuse channels derived from the photometric acquisition system. Consequently, the intermediate fine-tuning stage may have biased the encoder toward feature statistics that were not fully compatible with the representation adopted in the proposed pipeline.

A second possible factor concerns the differences in annotation strategy and defect morphology between the two datasets. The Bristol study considered separate segmentation tasks for surface defects and foreign objects, whereas the proposed dataset focuses on a different set of industrial defect structures and sensor-derived representations. As a result, the learned feature representations may not transfer efficiently between the two tasks despite their shared industrial context. More generally, these results suggest that representation compatibility and acquisition characteristics may play a more important role than industrial-domain similarity alone in cross-dataset transfer-learning scenarios for carbon-fibre defect segmentation. This observation further reinforces the importance of the representation-design analysis discussed throughout this thesis.

6.5 Discussion of Failure Modes and Remaining Limitations

Although the investigated methodological choices improve segmentation performance, several limitations remain unresolved.

6.5.1 Data scarcity

The primary limitation remains the scarcity of truly independent defect instances. Even after introducing improved representations, augmentation strategies, and leakage-safe training methodologies, realistic generalization remains substantially more difficult than initially suggested by optimistic random-splitting experiments.

This observation reinforces one of the central conclusions of this thesis: reliable evaluation protocols are fundamentally important in industrial inspection scenarios involving grouped acquisitions and limited datasets.

6.5.2 Material variability

Defect appearance depends strongly on material texture, fibre orientation, reflectance behaviour, and acquisition geometry. As a consequence, generalization to previously unseen materials remains particularly challenging.

The experiments suggest that the investigated models may still partially rely on material-specific texture statistics despite the adopted normalization strategies.

6.5.3 Small and low-contrast defects

Certain defect categories remain difficult to detect reliably, particularly:

- extremely small defects;
- low-contrast anomalies;
- orientation-related structural irregularities.

These cases often produce uncertain or incomplete predictions even for the best-performing configurations.

6.5.4 Limits of augmentation

Although CutMix augmentation improves segmentation performance, it does not fully solve the scarcity problem. Synthetic combinations cannot perfectly reproduce the variability and structural consistency of real industrial defects and may occasionally introduce unrealistic texture transitions. This suggests that augmentation should primarily be interpreted as a robustness-improvement strategy rather than as a replacement for genuinely diverse industrial training data.

6.5.5 Overall interpretation

Taken together, the experiments presented throughout this chapter demonstrate that industrial carbon-fibre defect segmentation remains strongly constrained by data scarcity, material variability, and the presence of correlated multi-acquisition observations.

The analysis showed that representation design, normalization strategies, backbone capacity, and leakage-safe augmentation can all improve segmentation robustness under realistic industrial conditions. However, the experiments also revealed that enforcing grouped leakage-safe evaluation substantially increases the apparent difficulty of the task compared with conventional random sample-level splitting protocols.

These observations suggest that optimistic evaluation methodologies may significantly overestimate real industrial generalization capability when correlated acquisitions are present. Consequently, one of the main outcomes of this work is not simply the identification of an effective segmentation pipeline, but a more realistic characterization of the challenges associated with industrial defect segmentation under leakage-safe evaluation conditions.

7 Conclusions and Future Work

This thesis investigated semantic segmentation for industrial carbon-fibre defect inspection under realistic data-scarcity conditions. Rather than focusing exclusively on maximizing segmentation performance, the work aimed to analyse the methodological factors that most strongly influence generalization under leakage-safe industrial evaluation conditions. One of the central observations of the thesis is that the apparent difficulty of the segmentation problem changes substantially once correlated acquisitions of the same physical defect instance are prevented from appearing simultaneously in the training and validation subsets. Under conventional random sample-level splitting strategies, segmentation performance appeared significantly more optimistic. However, enforcing grouped leakage-safe evaluation revealed a much more challenging and realistic industrial generalization scenario. This observation highlights an important issue for industrial computer vision: evaluation methodology can influence the apparent performance of a system as strongly as the architecture itself. When grouped acquisitions are present, overly permissive evaluation protocols may substantially overestimate real deployment capability.

Within this more rigorous evaluation setting, the thesis analysed several methodological components of the segmentation pipeline, including representation design, normalization strategies, frequency-domain preprocessing, backbone capacity, augmentation, and transfer-learning configurations. The experiments showed that representation design plays a major role when working with physics-derived industrial sensing data. Different sensor-derived channels encode complementary information related to fibre orientation, reflectance behaviour, and structural surface properties, and the effectiveness of a representation depends strongly on the considered defect morphology and material configuration. The analysis additionally showed that careful augmentation and normalization strategies can improve segmentation robustness under severe data scarcity conditions. However, the experiments also demonstrated that these improvements remain fundamentally constrained by the limited number of truly independent defect instances available for training. More broadly, the results suggest that industrial defect segmentation cannot be interpreted purely as an architecture-design problem. Under realistic industrial conditions, dataset structure, acquisition methodology, annotation consis-

tency, and evaluation protocol design become equally important components of the learning pipeline. Consequently, one of the primary outcomes of this work is not simply the identification of a specific segmentation configuration, but a more realistic characterization of the challenges associated with industrial defect segmentation under leakage-safe evaluation conditions.

Several directions for future work emerge naturally from the present analysis. The most important future direction concerns the acquisition of larger and more diverse industrial datasets containing additional independent defect instances, material types, and acquisition conditions. The experiments strongly suggest that data diversity currently represents the primary limiting factor for robust industrial generalization. Another promising direction concerns physically-aware augmentation strategies capable of generating more realistic synthetic defect configurations while preserving fibre-orientation consistency and material compatibility. Future work could additionally investigate hybrid inspection frameworks combining supervised segmentation and anomaly-detection approaches in order to improve robustness to previously unseen defect categories. Finally, an important long-term objective concerns the deployment of robust segmentation systems under continuously evolving industrial conditions, including material drift, acquisition variability, and changing defect distributions over time.

In conclusion, this thesis demonstrated that industrial carbon-fibre defect segmentation remains a challenging open problem once realistic leakage-safe evaluation conditions are enforced. At the same time, the work showed that rigorous evaluation methodology, representation design, and carefully constructed training pipelines can substantially improve robustness under severe industrial data scarcity conditions.

Rather than presenting the problem as fully solved, the thesis provides a more realistic and experimentally grounded analysis of the factors currently limiting industrial generalization, thereby establishing a stronger methodological foundation for future research in industrial defect segmentation.

Bibliography

- [1] E. Xie, W. Wang, Z. Yu, A. Anandkumar, J. M. Alvarez, and P. Luo, *Segformer: Simple and efficient design for semantic segmentation with transformers*, 2021. arXiv: 2105.15203 [cs.CV]. [Online]. Available: <https://arxiv.org/abs/2105.15203>
- [2] G. Jocher, J. Qiu, M. Liu, S. Lyu, F. C. Akyon, and M. E. Kalfaoglu, *Ultralytics yolo26: Unified real-time end-to-end vision models*, 2026. arXiv: 2606.03748 [cs.CV]. [Online]. Available: <https://arxiv.org/abs/2606.03748>
- [3] J. Silvestre-Blanes, T. Albero-Albero, I. Miralles, R. Pérez-Llorens, and J. Moreno, “A public fabric database for defect detection methods and results,” *Autex Research Journal*, vol. 19, Jun. 2019. DOI: 10.2478/aut-2019-0035
- [4] H. Xu, C. Liu, S. Duan, L. Ren, G. Cheng, and B. Hao, “A fabric defect segmentation model based on improved swin-unet with gabor filter,” *Applied Sciences*, vol. 13, no. 20, 2023, ISSN: 2076-3417. DOI: 10.3390/app132011386 [Online]. Available: <https://www.mdpi.com/2076-3417/13/20/11386>
- [5] A. P. Kandiyil, G. Atkinson, and M. Damghani, “Localisation of defects on carbon fibre surfaces using deep learning,” *NDT & E International*, vol. 157, p. 103498, 2026, ISSN: 0963-8695. DOI: 10.1016/j.ndteint.2025.103498 [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S0963869525001793>
- [6] A. Kolli, D. Krajnc, and C. Eitzinger, “Segmentation-aware attention mechanism for defect classification of both virgin and recycled carbon fiber fabric,” in *SAMPE Conference*, 2025. [Online]. Available: <https://www.compass-horizon.eu/wp-content/uploads/2025/11/Segmentation-Aware-Attention-Mechanism-for-Defect-Classification-of-both-Virgin-and-Recycled-Carbon-Fiber-Fabric.pdf>
- [7] PROFACTOR GmbH, *Industrial carbon-fibre defect dataset*, <https://zenodo.org/records/14966026>, 2025.
- [8] OpenCV, *Cvat: Computer vision annotation tool*, 2025. [Online]. Available: <https://github.com/opencv/cvat>

- [9] B. Zhou, H. Zhao, X. Puig, S. Fidler, A. Barriuso, and A. Torralba, “Scene parsing through ade20k dataset,” in *CVPR*, 2017.
- [10] S. Zambal, W. Palfinger, M. Stöger, and C. Eitzinger, “Accurate fibre orientation measurement for carbon fibre surfaces,” *Pattern Recognition*, vol. 48, no. 11, pp. 3324–3332, 2015, ISSN: 0031-3203. DOI: <https://doi.org/10.1016/j.patcog.2014.11.009> [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S0031320314004798>
- [11] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and L. Fei-Fei, “Imagenet: A large-scale hierarchical image database,” *CVPR*, 2009.
- [12] N. Abraham and N. M. Khan, “A novel focal tversky loss function with improved attention u-net for lesion segmentation,” *ISBI*, 2019.
- [13] S. S. M. Salehi, D. Erdogmus, and A. Gholipour, “Tversky loss function for image segmentation using 3d fully convolutional deep networks,” in *MLMI*, 2017.
- [14] I. Loshchilov and F. Hutter, “Sgdr: Stochastic gradient descent with warm restarts,” in *ICLR*, 2017.
- [15] L. N. Smith, “A disciplined approach to neural network hyper-parameters: Part 1 – learning rate, batch size, momentum, and weight decay,” *arXiv preprint arXiv:1803.09820*, 2018.
- [16] S. Yun, D. Han, S. J. Oh, S. Chun, J. Choe, and Y. Yoo, *Cutmix: Regularization strategy to train strong classifiers with localizable features*, 2019. arXiv: 1905.04899 [cs.CV]. [Online]. Available: <https://arxiv.org/abs/1905.04899>
- [17] K. Roth, L. Pemula, J. Zepeda, B. Schölkopf, T. Brox, and P. Gehler, *Towards total recall in industrial anomaly detection*, 2022. arXiv: 2106.08265 [cs.CV]. [Online]. Available: <https://arxiv.org/abs/2106.08265>