Data Science and Artificial Intelligence

Automatic Detection and Whole-Roll Localization of Deep Quilting in Rewinding Videos of Silicon Anode Material

Zhemin Xie

Supervisors: Dr. Lu Cao and Prof. Fons Verbeek
Company: LeydenJar Technologies B.V.

BACHELOR THESIS

Leiden Institute of Advanced Computer Science (LIACS)
www.liacs.leidenuniv.nl

19/06/2026

Abstract

LeydenJar Technologies B.V. develops pure silicon anode material for lithium-ion battery applications using roll-to-roll plasma-enhanced chemical vapour deposition (PECVD). During rewinding, the produced anode rolls may contain both normal visual texture variations and surface defects that can indicate downstream processing or yield risk. This thesis focuses on deep quilting, also referred to as tension wrinkles, which is treated as a production-relevant wrinkle-like surface deformation that should be detected, localized, and summarized for quality control review. The aim is therefore to develop and evaluate a computer vision pipeline that can identify deep quilting in rewinding videos and convert frame-level visual evidence into roll-level localization information.

The task is challenging for two reasons. First, deep quilting appears as an irregular localized structure on top of normal quilting background texture, making it difficult to separate from ordinary visual variation. Second, quality control requires roll-level localization rather than only frame-level detection. The proposed pipeline follows four main steps. First, frames are extracted from rewinding videos and cropped to a predefined central inspection area. Second, manual polygon annotations are used to train and evaluate detection and segmentation models. Third, node-aware candidate verification is explored as an additional structural cue. Finally, frame-level predictions are merged and mapped to approximate roll positions, producing meter-level event counts and density estimates for quality control review.

The main result is that YOLOv8s-seg provides the strongest and most stable segmentation performance among the tested models, outperforming the bounding-box baseline in strict localization quality. Node-aware verification provides useful structural evidence but does not consistently improve object-level precision while preserving recall. The final pipeline links the detection of a production-relevant visual defect to roll-level event counts and density statistics, making the output more suitable for quality control review and for distinguishing deep quilting candidates from normal visual texture during inspection.

Contents

1 Introduction

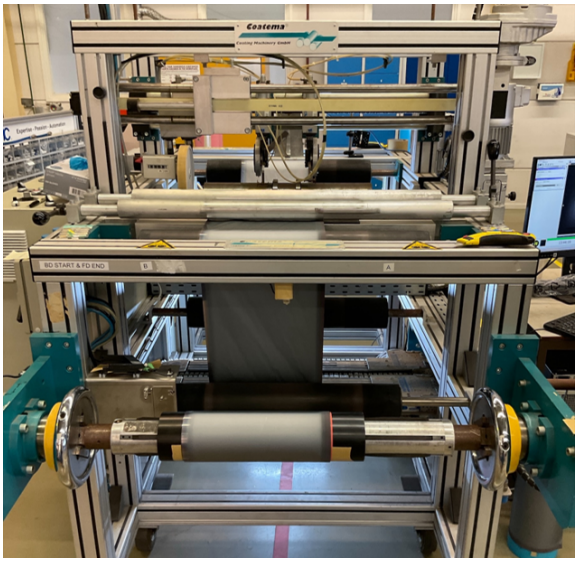
This chapter introduces the industrial and scientific context of the thesis. It explains the deep quilting defect, motivates the need for whole-roll localization in rewinding videos, defines the research aim and research questions, and summarizes the scope and contributions of the work.

1.1 Scientific and Industrial Context

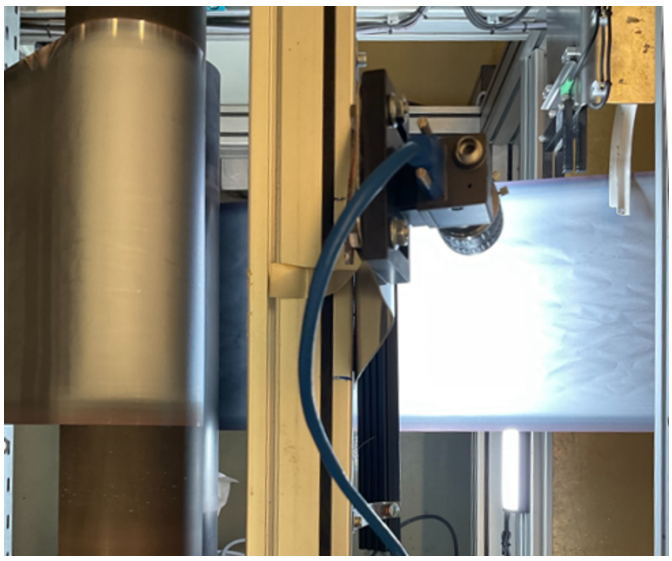
Automated visual inspection is an important component of quality control in industrial production, especially when materials are produced continuously and cannot be inspected reliably by manual observation alone. In the present roll-to-roll setting, the silicon anode material moves continuously through the camera field of view. The inspection system must therefore remain stable under motion, illumination variation, repeated surface texture, and production constraints. The task is not only to classify whether an image is normal or abnormal, but also to localize defect regions and provide information that can support quality control decisions [NJ95, MPZ+03, NMD14]. Deep learning has become a common approach for industrial surface defect detection [TŠSS20, LDJ+20]. Object detection models predict bounding boxes and confidence scores for candidate defects, while segmentation models assign individual pixels to defect regions and therefore produce pixel-level masks of the affected area [RHGS15]. This distinction is important for irregular surface defects, because a bounding box may indicate where a defect is located, but it does not describe its shape or area accurately. Segmentation can therefore provide more useful information for estimating defect extent, spatial distribution, and severity [LSD15, HGDG17]. This thesis is positioned at the intersection of automated visual inspection, surface defect segmentation, and roll-level quality control for continuous silicon anode material.

1.2 Practical Problem: Deep Quilting in Rewinding Videos

LeydenJar Technologies B.V. produces pure-silicon anode material using roll-to-roll plasma-enhanced chemical vapour deposition (PECVD). PECVD is a vacuum deposition process in which plasma is used to drive chemical reactions and deposit a thin film onto a substrate; in this application, the process is used to deposit silicon material in a continuous roll-to-roll setting. In this production context, thermal conditions during deposition and tension during material transport can be associated with quilting-type surface patterns. During production and rewinding, a regular fish-net-like surface pattern can therefore appear on the material. This normal quilting pattern is usually present across the roll and is treated as background rather than as the defect of interest. However, some abnormal patterns contain stronger localized wrinkle-like surface deformation. In this thesis, these abnormal localized structures are referred to as deep quilting or tension wrinkles.
Deep quilting is relevant because it can interfere with downstream battery cell manufacturing. Local surface deformation may reduce the usability of parts of the produced material and can contribute to yield reductions during battery cell production. Therefore, the industrial goal is not only to visually detect deep quilting but also to quantify where it occurs and how strongly it appears along the roll.



(a) Overview of the rewinding setup.



(b) Close-up of the camera view and illuminated material surface.

Figure 1: Industrial rewinding setup used for video acquisition. Panel (a) shows the rewinding setup, while panel (b) shows the camera view and illuminated material surface. The figure illustrates the roll-to-roll inspection context in which the silicon anode material moves through the camera field of view during rewinding.

1.3 Research Gap and Scientific Problem

Existing work on industrial surface defect detection often focuses on detecting defects in individual images or local image regions [LDJ+20, TŠSS20]. In contrast, the practical problem in this thesis also requires whole-roll localization. As illustrated by the rewinding setup in Figure 1, the camera records a continuously moving material sheet rather than isolated static samples. The output of the model must therefore be linked back to the position of the material along the roll. This turns the task into an inspection pipeline: videos have to be sampled into frames, a central inspection area has to be cropped, deep quilting has to be annotated, detection or segmentation models have to be trained, and frame-level predictions have to be merged and summarized at roll level. This final aggregation step follows the broader idea that overlapping visual observations can be aligned and combined into a consistent output [Sze07].

A second challenge is the visual similarity between acceptable background texture and target defects. Normal quilting may form a repeated surface pattern, while deep quilting appears as a more pronounced and localized structure. As a result, the problem cannot be reduced to a simple image-level distinction between normal and abnormal frames. A single frame may contain both normal quilting background and a local deep quilting structure, and the system must identify the target structure rather than the general texture of the material.

A third challenge concerns evaluation. Rewinding videos produce temporally adjacent frames that are visually similar. If frames from the same roll were randomly assigned to training and test sets, near-duplicate visual patterns could appear in both sets. This would create a data leakage risk and could overestimate model performance. To reduce this risk, this thesis uses roll-based evaluation, where the held-out test data come from a different roll than the training data. In addition, because

the material is moving during acquisition, timestamp-to-position mapping may be affected by motion-related temporal or geometric effects [LCC08, FR10].
Finally, the thesis investigates whether possible ways of improving the detection pipeline can be identified, for example by testing whether structural node information helps distinguish true deep quilting from texture-like false positives. For quality control use, recall is important because missing a relevant defect may leave affected material unflagged. However, precision also remains important, because too many false positives increase the amount of manual review. The evaluation therefore reports both precision and recall, with additional attention to the recall-oriented F2 score.

1.4 Research Aim and Research Questions

The aim of this thesis is to develop and evaluate a structured computer vision pipeline for detecting deep quilting in rewinding videos and converting frame-level detections into roll-level quality control statistics. The pipeline should support both image-level localization of deep quilting and roll-level summarization of detected events.
The main research question is:

> To what extent can a computer vision pipeline detect and localize deep quilting in rewinding videos of silicon anode material and summarize the detected defects at roll level for quality control?

This main question is divided into four sub-questions:

1. How can rewinding videos be processed and annotated to create a reliable dataset for deep quilting detection?
2. How well does segmentation-based detection perform compared with an earlier bounding-box detection baseline?
3. Does adding node information improve object-level detection performance while maintaining recall?
4. How can frame-level predictions be converted into roll-level statistics such as defect count and density per meter?

1.5 Research Approach and Scope

Data preparation and annotation. Frames were extracted from rewinding videos and cropped to a predefined central inspection area. This reduced irrelevant image regions and focused the dataset on the material area used for inspection. Deep quilting structures were then manually annotated with polygon labels in CVAT, providing ground-truth regions for model training and evaluation. Normal quilting was not annotated as a target class and was treated as background. To reduce leakage between visually similar frames, the dataset was split by roll and evaluated with a roll-based three-fold setup.

Modelling. The modelling part of the thesis was developed in stages. A bounding-box detection baseline was first used to test whether the deep quilting signal could be learned from the annotated images. The main model was then upgraded to a segmentation-based approach, because segmentation masks better represent irregular defect regions and can be used for density-based roll-level reporting [TŠSS20]. In addition, a node-aware extension was explored. This extension did not replace the segmentation model. Instead, the segmentation model first generated candidate regions, after which a node detector and candidate verifier were used to assess whether structural node evidence could improve object-level decisions.

Roll-level reporting. The final step was to convert frame-level predictions into roll-level reporting. Predicted masks were linked to frame indices or timestamps and mapped to approximate roll coordinates. Overlapping predictions within frames and across nearby frames were merged to reduce duplicate counting, following the same general principle of combining overlapping observations into a consistent output [Sze07]. The resulting report summarized deep quilting by meter-level event count and mask density, making the model output more directly useful for quality control review. The scope of this thesis is limited to detecting and localizing deep quilting or tension wrinkles. Normal quilting is treated as background. Other visual defects, such as pinholes, scratches, creases, and rainbow-like patterns, are outside the scope of this work because they may require different labels, imaging conditions, and evaluation criteria. Camera and lighting optimization were relevant to image quality but were not treated as the main research contribution.

1.6 Contributions and Thesis Structure

This thesis makes four main contributions. First, it defines a data preparation and annotation workflow for detecting deep quilting in rewinding videos. Second, it evaluates deep quilting detection with a roll-based split so that test images come from rolls not used for training. Third, it compares bounding-box detection, segmentation, and node-aware verification to determine which modelling approach is most suitable for this defect. Fourth, it converts frame-level predictions into roll-level count and density reports, which are more directly useful for quality control review.
The remainder of the thesis is structured as follows. Chapter 2 reviews related work on automated visual inspection, surface defect detection, segmentation, anomaly detection, and roll-level localization. Chapter 3 describes the data, annotation protocol, model design, roll-based splitting strategy, evaluation protocol, and roll-level reporting method. Chapter 4 presents the experimental results. Chapter 5 discusses the findings, answers the research questions, explains the practical implications, limitations, and future work, and concludes the thesis.

2 Background and Related Work

This chapter reviews the research areas that support the proposed inspection pipeline. It discusses automated visual inspection for continuous materials, the visual definition of deep quilting, detection and segmentation methods for surface defects, high-recall candidate generation, and video-based roll-position mapping.

2.1 Roll-to-Roll Visual Inspection and Surface Defects

Automated visual inspection has long been used to support industrial quality control, especially in continuous production settings where manual inspection cannot reliably cover the full material length. Newman and Jain reviewed automated visual inspection systems and showed that such systems are typically designed according to the imaging setup, product characteristics, and inspection objective [NJ95]. Malamas et al. further reviewed industrial machine vision systems from an application perspective and emphasized the role of image acquisition, defect definition, detection method, and deployment context [MPZ+03]. These studies show that industrial visual inspection is a system-level problem rather than only a model selection problem.

For continuous materials, visual inspection must also handle material-strip motion, illumination stability, and repeated surface texture. Neogi reviewed vision-based steel surface inspection systems and identified motion conditions, complex backgrounds, defect diversity, and industrial processing requirements as important challenges [NMD14]. Tang similarly described machine-vision-based surface defect detection for steel products as a problem involving imaging conditions, detection algorithms, and deployment requirements [TCSL23]. Although this thesis focuses on roll-to-roll silicon anode material rather than steel strip, both settings involve continuous material surfaces and localized visual defects. These studies therefore provide a relevant methodological background for the present work. In this thesis, motion and illumination effects are treated as acquisition and validity constraints rather than as separate camera-optimization targets. They motivate the use of a fixed imaging setup, a predefined central inspection crop, roll-based evaluation, and later error analysis.

This thesis builds on this background but adds a roll-level quality control requirement. The model output must not only indicate whether deep quilting is present in a cropped inspection image but also be converted into position and summary information along the full roll. The work therefore links automated visual inspection with whole-roll localization and quality control reporting.

2.2 Deep Quilting as a Visual Defect

The target defect in this thesis is deep quilting. It must be distinguished from normal quilting. Normal quilting is a more uniformly distributed fish-net-like surface pattern and is treated as background in this study. Deep quilting, also referred to as tension wrinkles, appears as a more localized and pronounced surface structure and often has a bamboo-like nodal appearance. Figure 2 illustrates this visual distinction. Since normal quilting and deep quilting can appear in the same cropped inspection image, the task cannot be reduced to frame-level normal or abnormal classification.

Visually, deep quilting is closer to an irregular surface defect than to an object with a fixed shape. It can be elongated, curved, partially visible, and variable across consecutive frames. A bounding

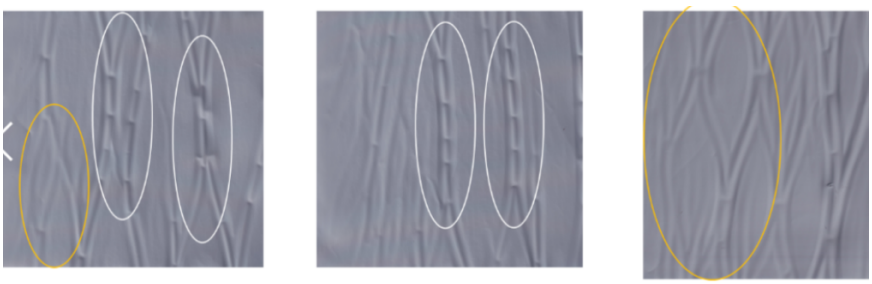


Figure 2: Visual comparison between normal quilting background texture and localized deep quilting structures. The yellow ellipses indicate normal quilting-like background texture, while the white ellipses indicate localized deep quilting or tension-wrinkle structures. The highlights are used only for visual explanation; model evaluation is based on manual polygon annotations rather than on these illustrative ellipses.

box can indicate its approximate location, but it loses part of the shape and area information. If the downstream task requires estimating defect extent or density, a pixel-level or region-level representation is more suitable.
Deep quilting is therefore not simply an image-level anomaly and not a standard object with a fixed shape. It is a localized irregular defect that appears on top of a complex background texture. The method must therefore address defect localization, shape representation, and roll-level aggregation.

2.3 Object Detection and Segmentation for Surface Defects

Industrial surface defect detection can be formulated as classification, object detection, or segmentation. Classification determines whether an image is abnormal, but it does not localize the defect. Object detection [RHGS15] provides bounding boxes and is therefore suitable for approximate localization. Segmentation produces pixel-level masks and is more suitable when defect shape, area, or spatial distribution is important [TŠSS20, LDJ+20].
Tabernik et al. proposed a segmentation-based deep-learning approach for surface-defect detection and evaluated it on real industrial surface defect data [TŠSS20]. Their work shows that segmentation-based methods can learn localized surface anomalies and that pixel-level defect regions are useful for practical inspection. Lv et al. introduced a metallic surface defect benchmark and detection network, illustrating that industrial surface defect detection often requires localization under complex texture and appearance conditions [LDJ+20].
These studies support the transition in this thesis from a bounding-box detection baseline to a segmentation-based model. Bounding-box detection can test whether the deep quilting signal is learnable, but it cannot fully represent the irregular shape of deep quilting. A segmentation model provides mask outputs that better describe the defect region and can be used directly for density-based roll-level reporting. For this reason, segmentation is the main modelling route in this thesis.

2.4 Anomaly Detection and High-Recall Candidate Generation

In industrial inspection, defect examples are often less common than normal examples, and defect appearance can vary. Data-efficient learning and high-recall candidate generation are therefore rele-

vant ideas in related work. For quality control, missing relevant defects is usually more problematic than producing additional candidates for review. Candidate generation can therefore be designed with recall in mind.

Roth et al. emphasized the value of high-recall anomaly localization in industrial anomaly detection [RPZ+22]. Although this thesis did not use an unsupervised anomaly detection dataset and did not adopt anomaly detection as the main method, the high-recall perspective remains methodologically relevant. In settings with complex defect appearance and high cost of missed defects, a system can first preserve a broader candidate set and then apply later decision steps.

This section therefore treats anomaly detection as related background for candidate generation, not as the main technical route of the thesis. The final method remains supervised defect detection and segmentation, and the final performance is evaluated using supervised detection metrics, object-level metrics, and roll-level reporting.

2.5 Video Sampling, Motion, and Roll-Position Mapping

The input data in this thesis come from rewinding videos, making frame sampling and roll-position mapping central parts of the method. Unlike static image inspection, consecutive video frames are temporally dependent and often visually similar. If frames from the same roll were randomly assigned to training and test sets, the evaluation could be overestimated because near-duplicate visual patterns might appear in both sets [RBC+17]. For this reason, this thesis uses roll-based splitting rather than random frame-level splitting.

Motion imaging can introduce temporal and geometric artifacts. For example, rolling-shutter acquisition may cause geometric distortion when the camera or object moves during exposure [LCC08, FR10]. More generally, the effective image quality depends on the relation between exposure time and material speed: if the material moves too far during exposure, motion blur or geometric uncertainty can make fine surface structures harder to interpret. In this thesis, these artifacts are treated as acquisition risks rather than as problems directly corrected by the proposed method. The imaging setup used a global shutter camera, and the exposure time was set to approximately 25,000 μs, or 25 ms. With the material speed treated as approximately 0.02 m/s, the material displacement during one exposure is about 0.5 mm. Therefore, the camera and acquisition settings were considered sufficient for reducing motion-related blur and geometric distortion in the current setup, although exposure optimization itself was not part of the proposed method. Instead, the method focuses on frame extraction, cropped inspection images, prediction merging, and roll-level reporting.

For whole-roll localization, frame indices or timestamps must be mapped to material-length coordinates. In this project, the material speed during rewinding was treated as approximately stable, so roll-position mapping was based on a constant-speed assumption. Detected deep quilting predictions also had to be merged across nearby frames to avoid counting the same structure multiple times. Szeliski's work on image alignment and stitching provides a broader methodological background for overlap-based merging [Sze07]. Although this thesis does not perform panoramic stitching, it uses the related idea of combining overlapping observations into consistent roll-level events.

3 Methods

This chapter describes how the inspection pipeline was implemented and evaluated. It covers the data source, imaging setup, central inspection crop, annotation protocol, roll-based splitting strategy, model experiments, node-aware verification, roll-level reporting, and evaluation protocol.

3.1 Methodological Overview

This study used a computer vision pipeline from rewinding videos to roll-level quality control reporting. As shown in Figure 3, the pipeline included frame extraction, central inspection crop generation, manual annotation, roll-based dataset splitting, model experiments, frame-level prediction merging, and final roll-level reporting. The following sections describe the data, annotation, model experiments, evaluation protocol, and roll-level reporting steps. Command-line calls, directory structure, and script-level notes were saved separately to support reproducibility.
The aim of the pipeline was not only to train a frame-level model but also to evaluate whether deep quilting could be localized in cropped inspection images and then converted into roll-level quality control statistics. The model experiments were organized into three stages. V1 was a YOLOv8s bounding-box detection baseline used to test whether the deep quilting signal was learnable with rectangular object localization. V2 was the main segmentation stage, where YOLOv8s-seg was trained and compared with Mask R-CNN and RTMDet-Ins. V3 was a node-aware verification stage built on top of V2 segmentation candidates. Its purpose was to test whether structural node evidence could help distinguish true deep quilting candidates from texture-like false positives.

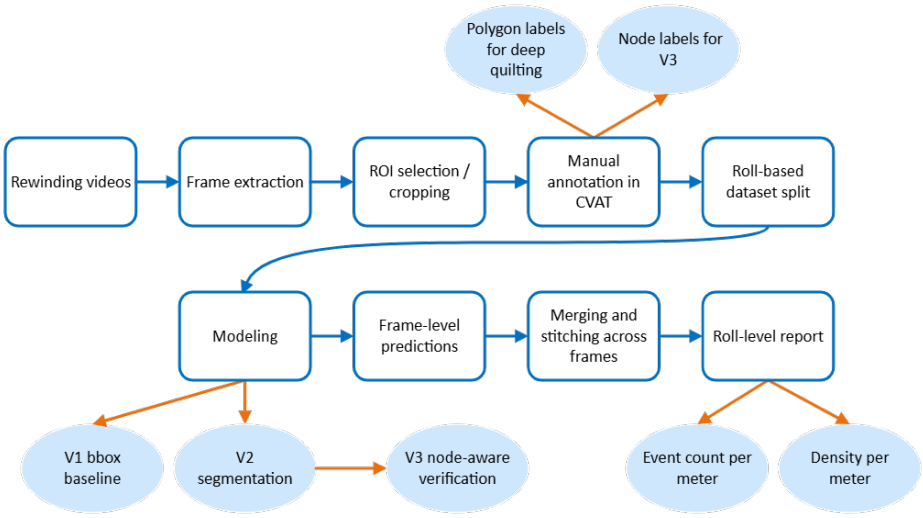


Figure 3: Overview of the research workflow and inspection pipeline from rewinding videos recorded at the rewinding setup of LeydenJar Technologies B.V. to roll-level quality control reporting. The workflow includes frame extraction, central inspection crop generation, manual annotation, roll-based dataset splitting, model experiments, frame-level prediction merging, and final roll-level reporting. The modelling stage is organized into three parts: V1 is a YOLOv8s bounding-box baseline, V2 is the main YOLOv8s-seg segmentation model, and V3 is a node-aware verification stage built on top of V2 segmentation candidates.

3.2 Data Source, Imaging Setup, and Central Inspection Crop

The input data consisted of rewinding videos recorded at the rewinding setup of LeydenJar Technologies B.V. This data-preparation stage corresponds to the first part of the research workflow shown in Figure 3. The videos were extracted into frames and then cropped to a predefined central inspection area for annotation, model training, and roll-level reporting. In this thesis, these cropped frame regions are referred to as cropped inspection images.

Cropped inspection images were used instead of full frames to remove edge regions, irrelevant background, and unstable image areas. The initial frame extraction rate was 1 frame per second. The material speed was treated as approximately 0.02 m/s. At this speed, two consecutive 1 fps frames corresponded to approximately 0.02 m, or 2 cm, of material travel. This frame extraction rate provided sufficient overlap between nearby frames for cross-frame merging while reducing storage, processing, and annotation workload.

The camera configuration and acquisition settings were selected to reduce motion-related geometric distortion. In particular, a global shutter configuration was used to avoid rolling-shutter distortion. The exposure time was set to approximately 25,000 μs, or 25 ms. With the material speed treated as 0.02 m/s, the material displacement during one exposure is approximately 0.5 mm. Therefore, motion imaging was mainly relevant to frame-to-roll mapping and repeated observations across nearby frames, rather than as a major image distortion problem requiring correction.

The central inspection crop was defined using a grid-based reference system. The full material-strip width was approximately 400 mm, and approximately 50 mm was excluded on each side. These edge regions were excluded because they can contain boundary effects, non-representative material areas, and less stable imaging conditions. The target of this thesis was the central material area used for deep-quilting inspection, so the main inspection area focused on the central 300 mm of the material strip. The practical crop used in this study corresponded to columns F-AD and rows 6-16. The F-AD range corresponded to approximately 303 mm and was used for pixel-to-millimetre conversion:

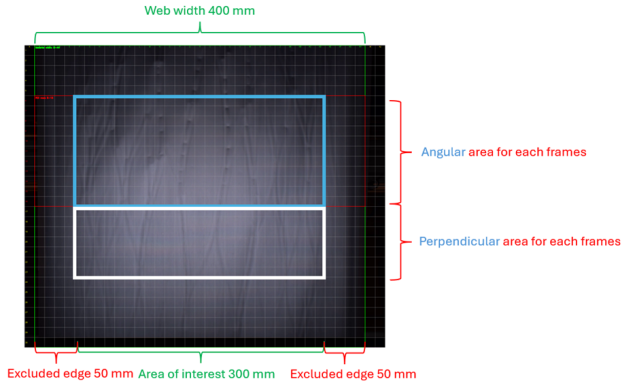
$$r_{\text{mm/px}} = \frac{W_{\text{crop, mm}}}{W_{\text{crop, px}}}, \tag{1}$$

where $W_{\text{crop, mm}} = 303$ mm. The central inspection crop definition and the visual definition of deep quilting nodes are shown in Figure 4.

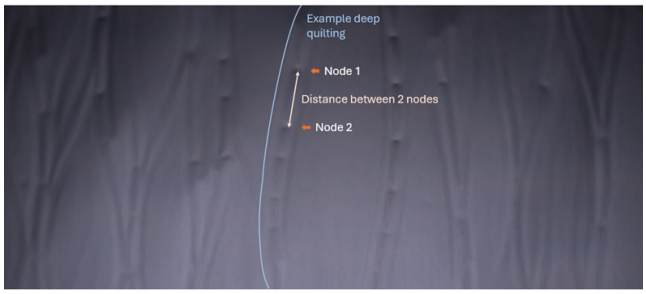
3.3 Deep Quilting Definition and Annotation Protocol

The target defect in this study was deep quilting, also referred to as tension wrinkles. Normal quilting was treated as a background texture and was not used as a target class. This distinction was necessary because normal quilting can appear as a repeated fish-net-like visual pattern, while deep quilting is a more localized and pronounced wrinkle-like structure that is relevant for quality control review. The definition used in this thesis is therefore an operational annotation definition: it was designed to consistently identify the target deep-quilting structures in cropped inspection images, rather than to classify all possible visual defects or to model the physical cause of yield loss. Deep quilting was defined as a localized structure with a bamboo-like nodal appearance inside the cropped inspection image. To improve annotation consistency, a candidate deep-quilting structure normally had to satisfy the following conditions:

- At least four nodes had to appear on the same continuous line-like structure.



(a) Central inspection crop on the material surface.



(b) Example of deep quilting nodes and node distance.

Figure 4: Operational definitions of the central inspection crop and deep quilting structure. Panel (a) shows the crop selection on the material surface. The edge regions are excluded from the target inspection area because the study focuses on the central material region used for deep-quilting analysis. Panel (b) shows an example deep quilting structure with connected nodes used for the annotation definition.

- The physical distance between two adjacent nodes should not exceed 34 mm.

The distance threshold was used to avoid merging visually disconnected nodes into one deep-quilting structure. These criteria helped separate localized deep-quilting structures from ordinary background quilting texture during manual annotation.
Manual annotation was performed in CVAT, an interactive video and image annotation tool for computer vision [CVA23]. Deep quilting was annotated as polygon objects because its boundary was irregular and because the spatial extent of the defect was required for segmentation and density estimation. These polygon annotations served as the ground-truth regions for model training and evaluation. One continuous deep-quilting structure was annotated as one object. Images without deep quilting were retained as negative samples. The V2 model used only deep-quilting polygons as its segmentation target, while node labels were retained as auxiliary structural information for V3.

3.4 Dataset Stages and Roll-Based Splitting

The experiments were organized into three data stages rather than three fully independent datasets. The first stage was the V1 bounding-box detection baseline, which used an early annotation set to test whether the deep quilting signal was learnable with object detection. The second stage was the V2 segmentation stage, which used merged polygon annotations as the main dataset for segmentation training and evaluation. The third stage was the V3 node-aware verification stage, which reused the V2 cropped inspection images, metadata, annotations, and roll-based split, but additionally used node labels and V2 candidate-level features for verification.
The baseline dataset came from the first annotation stage and contained 400 cropped inspection images and 190 annotations. Both positive and negative images were present. The practice roll contained a substantially higher proportion of positive deep quilting examples than the normal rolls. For this reason, it was not included as one of the three held-out test rolls. Instead, it was used only as a training-only supplementary roll. This allowed the training set to include additional

positive examples without changing the roll-based evaluation protocol or using the practice roll for validation or testing.

The V2 dataset was built by merging two CVAT annotation batches. The final merged dataset contained 1250 images and 2335 annotations. The main rolls were Roll1, Roll2, and Roll3, while the practice roll remained training augmentation. The V2 annotation files retained both deep quilting polygons and node labels, but only the deep quilting polygon labels were used for the main V2 segmentation task. The node labels were retained for V3 because deep quilting often has a bamboo-like nodal structure. A candidate that contains consistent node evidence may be more likely to represent true deep quilting, while texture-like false positives may lack this structure. V3 therefore tested whether node information could provide auxiliary structural evidence for candidate verification.

The V3 dataset inherited the V2 images, annotations, metadata, and fold split. It additionally linked deep quilting instances with node labels. This produced 426 deep quilting instances and 1909 assigned nodes. Positive node crops, background negative crops, and V2 false-positive hard negative crops were then generated for node detector training and candidate-level feature extraction.

All main model experiments used roll-based three-fold splitting instead of random frame-level splitting. This was done to reduce leakage caused by visually similar neighboring frames from the same roll. For data with temporal, spatial, or hierarchical dependence, random cross-validation can underestimate prediction error, and block-based validation strategies are recommended [RBC+17]. The fold assignment is shown in Table 1.

Table 1: Roll-based three-fold split used for model evaluation. The practice roll was used only as supplementary training data and was not included in validation or held-out testing.

Fold	Test roll	Main training rolls	Training-only supplementary roll
Fold 1	Roll1	Roll2, Roll3	practice roll
Fold 2	Roll2	Roll1, Roll3	practice roll
Fold 3	Roll3	Roll1, Roll2	practice roll

3.5 Detection and Segmentation Model Experiments

The first model was a bounding-box detection baseline. The polygon annotations were converted into bounding boxes and used to train a YOLOv8s detection model. YOLOv8 is an Ultralytics model family that supports object detection, instance segmentation, classification, and related computer vision tasks [JCQ23]. The purpose of the baseline was to test whether the deep quilting signal was learnable from cropped inspection images, not to provide the final quality control model.

The main V2 model was YOLOv8s-seg instance segmentation. Segmentation was selected because deep quilting is an irregular, elongated, and locally variable defect structure. A mask represents the defect region more appropriately than a bounding box and can be used directly for density-based roll-level reporting. The primary V2 experiment used the roll-based three-fold setup with 80 epochs, image size 960, batch size 8, and 4 workers.

To assess whether the segmentation result depended strongly on a single architecture, two additional instance segmentation models were included as practical baselines: Mask R-CNN and RTMDet-Ins. Mask R-CNN is a standard instance segmentation framework that adds a mask prediction branch to

an object detection framework [HGDG17]. RTMDet was designed as an efficient real-time detector and can be extended to instance segmentation tasks [LZH+22]. For comparability, these models used the same roll-based data split and the same training, validation, and test data sources as the YOLO segmentation experiments.

After the primary three-fold V2 YOLOv8s-seg segmentation experiment, the selected Fold 3 V2 YOLOv8s-seg model was used for a secondary tuning analysis. This step examined whether changing selected training settings, such as the initial learning rate, final learning-rate factor, weight decay, and batch size, could improve the precision-recall balance on a selected representative fold. It was not treated as a replacement for the primary three-fold evaluation, and the final comparison with Mask R-CNN and RTMDet-Ins remained based on the original, untuned three-fold YOLOv8s-seg results.

The V3 node-aware experiment was then introduced as a separate candidate-verification analysis. Its purpose was not to train another segmentation model, but to test whether structural node evidence could help filter or rank V2 segmentation candidates.

3.6 V3 Node-Aware Verification of V2 Candidates

V3 was designed as an exploratory verification stage built on top of the V2 segmentation model. It did not replace V2 segmentation. Instead, V2 first generated candidate deep quilting regions from cropped inspection images. V3 then tested whether additional node information could help decide whether each V2 candidate should be kept as a likely deep quilting instance or treated as a likely false positive.

The motivation for V3 comes from the visual structure of deep quilting. True deep quilting often contains a bamboo-like nodal pattern, where several connected nodes appear along the same elongated structure. Some false positives from V2, however, are caused by background texture or local surface patterns that resemble deep quilting but do not contain consistent node evidence. Node-aware verification was therefore introduced to test whether this structural cue could improve candidate-level precision while maintaining recall.

The V3 workflow consisted of four steps. First, the V2 YOLOv8s-seg model generated segmentation candidates at a selected operating point. Second, each candidate region was cropped and passed to a YOLOv8n node detector trained on node annotations. Third, candidate-level features were extracted from both the V2 segmentation output and the node detector output. These features included segmentation confidence, mask area, bounding-box geometry, node count, average node confidence, and node density inside the candidate region. Fourth, a lightweight logistic regression verifier used these features to classify each candidate as likely true deep quilting or likely false positive.

Three practical candidate-handling strategies were compared. The first strategy, *V2 accept-all candidates*, kept all V2 segmentation candidates and served as the recall-oriented candidate baseline. The second strategy, *node-feature logistic verifier*, used a logistic regression verifier based on segmentation and node features to accept or reject V2 candidates. The third strategy, *review-oriented node-assisted policy*, assigned candidates to accepted, review-needed, and rejected groups. For evaluation, both accepted and review-needed candidates were treated as surfaced system outputs because they would still be presented for human review in a quality control workflow.

V3 was evaluated at the object level. Each surfaced candidate was matched against ground-truth deep quilting instances and counted as TP_{obj}, FP_{obj}, or FN_{obj}. The reported metrics were

object-level precision, object-level recall, F1, and F2. The main question was whether node-aware verification could reduce false positives from V2 candidates without substantially increasing false negatives.

3.7 Roll-Level Localization and Reporting

Frame-level predictions were converted into roll-level reports. Since the material speed during rewinding was approximately stable, a constant-speed mapping was used. If T_m is the time required for the material to move one meter and Δt is the time interval between adjacent frames, the material displacement between the frames is:

$$\Delta s = \frac{\Delta t}{T_m}. \tag{2}$$

For timestamp t, the roll coordinate was approximated as:

$$S(t) = \frac{t}{T_m}. \tag{3}$$

The general form $S(t) = \int_0^t v(\tau)d\tau$ was not used as the main mapping method because the material speed was considered sufficiently stable for this study.

Overlapping predictions within the same frame were first merged to avoid representing the same deep quilting structure with multiple masks. Predictions in nearby frames were then stitched into roll-level events when their roll-coordinate intervals and spatial positions were consistent. This step follows the general principle of merging overlapping visual observations into a consistent output, which is also central to image alignment and stitching methods [Sze07].

The final roll-level report was calculated using meter bins. Two summary measures were produced: event count per meter and density per meter. Meter-level density was defined as follows:

$$D_m = \frac{\sum_{i \in F_m} A_i^{\text{mask}}}{\sum_{i \in F_m} A_i^{\text{crop}}} \times 100\%. \tag{4}$$

Here, F_m is the set of frames assigned to meter bin m, A_i^{mask} is the predicted mask union area in frame i, and A_i^{crop} is the inspected crop area in that frame.

3.8 Evaluation Protocol and Reproducibility

The evaluation was organized at three levels: model-level evaluation, object-level evaluation, and roll-level evaluation. These levels correspond to different stages of the inspection pipeline. Where ground-truth comparison was required, performance was measured with respect to the manual deep-quilting polygon annotations described in Section 3.3.

Model-level evaluation measured the direct output of trained detection or segmentation models on held-out cropped inspection images. This level was used for the bounding-box baseline, YOLOv8s-seg, Mask R-CNN, and RTMDet-Ins. The reported metrics included precision, recall, mAP50, and mAP50-95. For segmentation models, both box-level and mask-level metrics were recorded, while mask-level metrics were treated as the main segmentation results.

Object-level evaluation was performed after model predictions had been converted into candidate defect instances at a selected operating point. Each predicted candidate was matched against the ground-truth deep-quilting instances and counted as a true positive, false positive, or false negative. This level was used for the V3 node-aware verification comparison and for the segmentation-model object-level comparison. The reported metrics included precision, recall, F1, and F2. F2 was included because recall is especially important in this quality control setting, where missed defects may leave affected material unflagged.

Roll-level evaluation assessed the final output after frame-level predictions had been merged across nearby frames and mapped to material-length coordinates. At this level, the purpose was not only to evaluate individual detections, but also to summarize where predicted deep quilting occurred along the roll. The reported roll-level outputs included event count per meter, density per meter, and high-risk meter intervals.

The implementation used Python, OpenCV [Bra00], CVAT [CVA23], Ultralytics YOLO [JCQ23], MMDetection [CWP+19], scikit-learn [PVG+11], and CUDA-enabled PyTorch [PGM+19]. Dataset versions, split files, model configurations, prediction outputs, and requirements files were saved to support reproducibility. Environment setup, command-line calls, directory structure, and script-level notes were documented separately from the main thesis text.

4 Results

This chapter presents the experimental results of the thesis. It reports the dataset composition, V1 bounding-box baseline results, V2 segmentation results, model comparison and YOLO tuning, error analysis, V3 node verification results, and roll-level reporting outputs.

4.1 Dataset and Annotation Summary

This section summarizes the final data composition used in the experiments, including image counts, annotation counts, and roll-level class distribution. The data-preparation workflow, annotation protocol, and roll-based splitting procedure were described in the Methods chapter. The experiments used three dataset versions: a baseline detection dataset, a V2 segmentation dataset, and a V3 node-aware dataset. The baseline detection dataset was built from the first annotation stage and contained 400 cropped inspection images and 190 annotations, with both positive and negative samples retained. The V2 segmentation dataset was constructed by merging two CVAT annotation batches and contained 1250 cropped inspection images, 426 deep quilting instances, and 1909 node annotations. The V3 node-aware dataset inherited the V2 images, metadata, and roll-based split, and further used the link between deep quilting instances and node labels for node-aware verification. A positive image is defined as a cropped inspection image that contains at least one manual deep-quilting polygon annotation. A negative image contains no manual deep-quilting polygon annotation and was retained as a background sample. Node annotations were used as auxiliary structural information for V3, but they did not determine whether an image was counted as positive or negative. Table 2 summarizes the scale and purpose of the three dataset versions. Table 3 reports the roll-level distribution of the V2 segmentation dataset. The three main rolls were Roll1, Roll2, and Roll3. Together, they contained 1150 images, including 231 positive images and 919 negative images. The practice roll contained 100 images, of which 89 were positive. Because its positive ratio was much higher than the main rolls, the practice roll was used only as training-only supplementary data and was not included in validation or test sets.

Table 2: Summary of dataset versions used in the experiments.

Dataset version	Main purpose	Images	Annotations / instances	Additional information
Baseline detection	Bounding-box baseline	400	190 annotations	Positive and negative images retained
V2 segmentation	Main segmentation experiments	1250	426 DQ instances; 1909 node annotations	Deep quilting polygons used for training
V3 node-aware	Candidate verification	1250	426 DQ instances; 1909 assigned nodes	Inherited V2 images, metadata, and split

Table 3: Roll-level summary of the V2 segmentation dataset.

Roll	Images	Positive images	Negative images	DQ instances	Nodes
Roll1	400	48	352	51	216
Roll2	400	92	308	104	438
Roll3	350	91	259	116	480
practice	100	89	11	155	775
Total	1250	320	930	426	1909

4.2 Baseline Bounding-Box Detection Results

The bounding-box detection baseline was evaluated under the roll-based three-fold setup. The purpose of the baseline was not to provide the final quality control model, but to establish an initial reference point for testing whether the deep quilting signal could be learned from cropped inspection images. It also provided a comparison point for the later segmentation-based method.
Table 4 reports the results across the three folds. For each fold, the reported epoch was selected by the highest value of `metrics/mAP50-95(B)`, because the baseline was a bounding-box detection model. Fold 1 achieved the highest mAP50-95(B), with a precision of 0.779, recall of 0.500, mAP50(B) of 0.662, and mAP50-95(B) of 0.403. Fold 2 achieved the highest recall, 0.750, but had a lower precision of 0.428. Fold 3 achieved the highest precision, 1.000, with a recall of 0.472. Across the three folds, the mean precision, recall, mAP50, and mAP50-95 were 0.736, 0.574, 0.593, and 0.286, respectively.
Overall, the baseline detector learned a meaningful deep quilting signal, but its performance varied across folds. The limited recall and mAP50-95 values indicate that bounding-box detection was not sufficient to represent the irregular structure of deep quilting. Therefore, the baseline was used as a reference point for the V2 segmentation model.

Table 4: Bounding-box detection baseline results across the roll-based three-fold evaluation. The reported epoch is selected by the highest box mAP50-95 in the training results.

Fold	Epoch	P(B)	R(B)	mAP50(B)	mAP50-95(B)
Fold 1	26	0.779	0.500	0.662	0.403
Fold 2	78	0.428	0.750	0.524	0.216
Fold 3	67	1.000	0.472	0.594	0.239
Mean	–	0.736	0.574	0.593	0.286

4.3 V2 YOLOv8s-seg Segmentation Results

YOLOv8s-seg was evaluated as the main V2 segmentation model under the roll-based three-fold setup. Since V2 is a segmentation-based method, both box-level and mask-level metrics are reported. The box-level metrics provide a conservative reference to the earlier bounding-box baseline, while the mask-level metrics are the main evaluation results for the V2 segmentation model.
Table 5 reports the best mask-localization result for each fold, selected by the highest value of `metrics/mAP50-95(M)` in the training log. Fold 1 achieved a mask-level precision of 0.692, recall of 0.702, mAP50(M) of 0.782, and mAP50-95(M) of 0.547. Fold 2 achieved a mask-level precision of 0.556, recall of 0.640, mAP50(M) of 0.571, and mAP50-95(M) of 0.403. Fold 3 achieved a mask-level precision of 0.770, recall of 0.640, mAP50(M) of 0.734, and mAP50-95(M) of 0.494. Across the three folds, the mean mask-level precision, recall, mAP50, and mAP50-95 were 0.673, 0.660, 0.696, and 0.481, respectively.
Overall, V2 segmentation produced measurable mask-level localization performance across all three held-out rolls. Fold 1 and Fold 3 performed more strongly than Fold 2, particularly in mAP50(M) and mAP50-95(M). Figure 5 visualizes the main mask-level metrics across folds using a grouped bar chart. Compared with the bounding-box baseline, V2 achieved higher mAP50-95 values across all three folds. As shown in Figure 6, the improvement was largest in Fold 2 and Fold 3, where the

stricter localization metric increased from 0.216 to 0.403 and from 0.239 to 0.494, respectively. This comparison supports the use of segmentation as the main V2 method.

Table 5: V2 YOLOv8s-seg segmentation results across the roll-based three-fold evaluation. The reported epoch is selected by the highest mask mAP50-95 in the training results.

Fold	Epoch	P(B)	R(B)	mAP50(B)	mAP50-95(B)	P(M)	R(M)	mAP50(M)	mAP50-95(M)
Fold 1	66	0.692	0.702	0.782	0.600	0.692	0.702	0.782	0.547
Fold 2	69	0.521	0.600	0.547	0.379	0.556	0.640	0.571	0.403
Fold 3	60	0.770	0.640	0.734	0.502	0.770	0.640	0.734	0.494
Mean	–	0.661	0.647	0.688	0.494	0.673	0.660	0.696	0.481

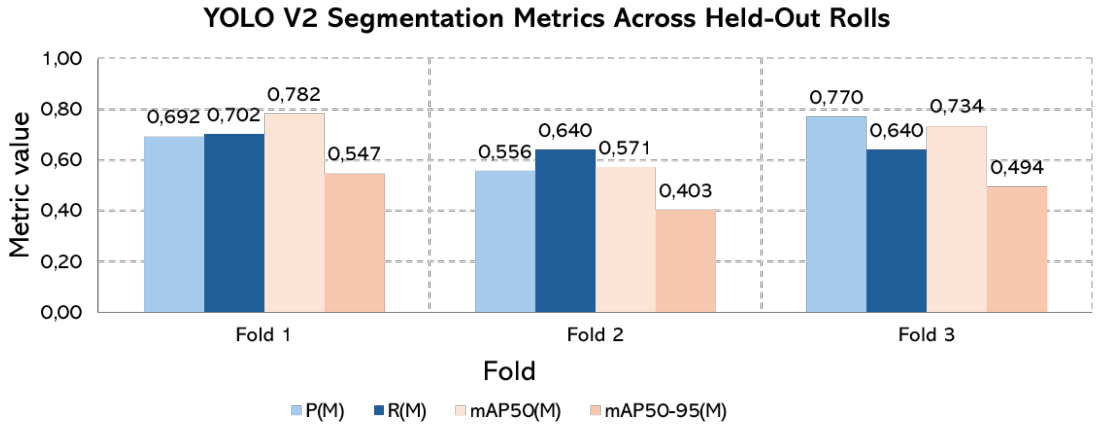


Figure 5: Mask-level performance of V2 YOLOv8s-seg across the three held-out rolls. The grouped bars show precision, recall, mAP50, and mAP50-95 for each fold.

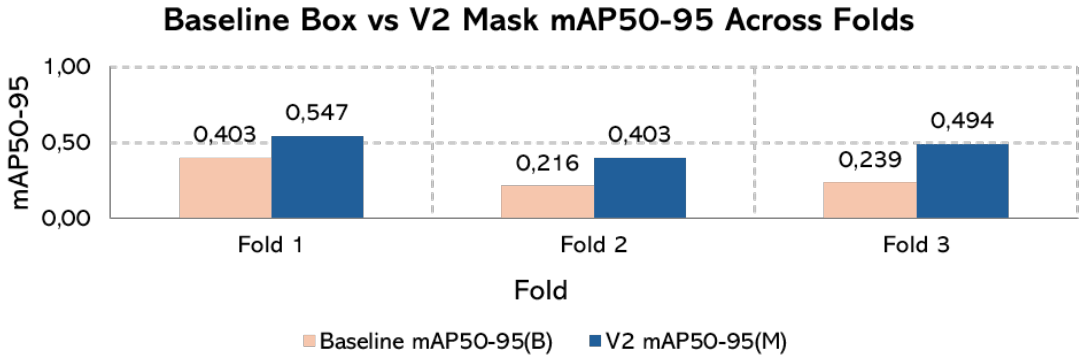


Figure 6: Comparison between the bounding-box baseline and the V2 segmentation model using strict localization metrics. The baseline is reported using box mAP50-95, while V2 is reported using mask mAP50-95.

4.4 Segmentation Model Comparison and YOLO Tuning

The V2 YOLOv8s-seg model, Mask R-CNN, and RTMDet-Ins were compared under the same roll-based evaluation setup. For this comparison, all three models were evaluated using their primary

three-fold results without the later YOLO tuning step. Mask R-CNN and RTMDet-Ins were used with their initial experimental settings as practical architecture baselines, rather than as extensively optimized models. The comparison therefore evaluates which model family was most suitable under the current data split and training setup, rather than the best possible performance after exhaustive hyperparameter search.

The comparison uses object-level TP, FP, FN, precision, recall, F1, and F2 instead of training-log mAP values, so that the outputs of the different model families can be compared at the candidate-decision level. The three models represent YOLO-based segmentation, a standard instance segmentation framework, and a real-time instance segmentation family, respectively.

Table 6 reports the object-level results for the three segmentation models across the three folds. YOLOv8s-seg produced non-zero recall and F2 in all three folds, with the highest F2 value of 0.704 on Fold 1. Mask R-CNN achieved high recall on Fold 1 and Fold 2, with recall values of 0.902 and 0.644, respectively, but produced many false positives on Fold 1 and an extremely large number of false-positive predictions on Fold 3. RTMDet-Ins did not produce true positive detections in any fold. Under the current data and training setup, YOLOv8s-seg was therefore the most stable segmentation model among the tested options.

Table 6: Object-level comparison of segmentation models across the roll-based folds.

Model	Fold	TP	FP	FN	Precision	Recall	F1	F2
YOLOv8s-seg	Fold 1	40	40	11	0.500	0.784	0.611	0.704
YOLOv8s-seg	Fold 2	62	44	42	0.585	0.596	0.590	0.594
YOLOv8s-seg	Fold 3	71	70	45	0.504	0.612	0.553	0.587
Mask R-CNN	Fold 1	46	126	5	0.267	0.902	0.413	0.612
Mask R-CNN	Fold 2	67	43	37	0.609	0.644	0.626	0.637
Mask R-CNN	Fold 3	0	35000	116	0.000	0.000	0.000	0.000
RTMDet-Ins	Fold 1	0	0	51	0.000	0.000	0.000	0.000
RTMDet-Ins	Fold 2	0	2	104	0.000	0.000	0.000	0.000
RTMDet-Ins	Fold 3	0	0	116	0.000	0.000	0.000	0.000

Figure 7 summarizes the model comparison using object-level F2, which is more recall-sensitive than F1.

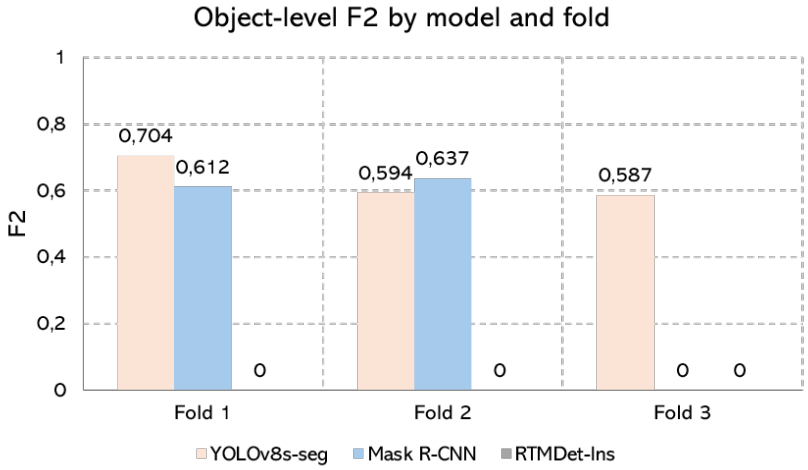


Figure 7: Object-level F2 comparison of YOLOv8s-seg, Mask R-CNN, and RTMDet-Ins across the roll-based folds. F2 is used because it weights recall more strongly than precision.

After the primary three-fold V2 YOLOv8s-seg experiment, a secondary tuning analysis was performed only on the selected Fold 3 V2 YOLOv8s-seg setup. Fold 3 was selected because it was one of the stronger folds in the primary segmentation experiment. This tuning step adjusted training settings such as the initial learning rate, final learning-rate factor, weight decay, and batch size. Because the tuned settings did not produce a clear and stable improvement over the default setting, the tuning analysis was reported only as a follow-up check and did not replace the primary roll-based evaluation. The final comparison with Mask R-CNN and RTMDet-Ins therefore used the original, untuned V2 YOLOv8s-seg results from the primary three-fold evaluation.

Table 7: YOLO tuning results on the selected Fold 3 model.

Experiment	Key setting	Epoch	P(M)	R(M)	mAP50(M)	mAP50-95(M)	Main change
Default Fold 3 YOLO	Default training setting	60	0.770	0.640	0.734	0.494	Reference
Best mAP50-95 tuned run	lr0=0.005, lrf=0.2, wd=0.0005, batch=8	74	0.704	0.566	0.704	0.499	+0.005 mAP50-95, lower recall
Recall-oriented selected run	lr0=0.005, lrf=0.2, wd=0.0003, batch=8	72	0.590	0.667	0.668	0.465	+0.027 recall, lower mAP50-95

The default Fold 3 YOLO model achieved a mask precision of 0.770, recall of 0.640, mAP50 of 0.734, and mAP50-95 of 0.494. The best mAP50-95 tuned run slightly increased mAP50-95 from 0.494 to 0.499, but recall decreased from 0.640 to 0.566. The recall-oriented selected run increased recall from 0.640 to 0.667, but mAP50-95 decreased to 0.465. Therefore, the follow-up tuning did not produce a setting that clearly improved precision, recall, and mask-localization quality at the same time.

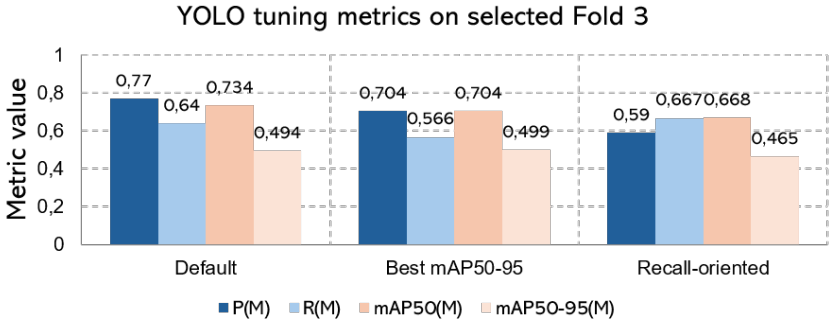


Figure 8: Comparison of the default and tuned training settings for the selected Fold 3 V2 YOLOv8s-seg model. The bars show mask-level precision, recall, mAP50, and mAP50-95.

4.5 Error Analysis of the Segmentation Model

This section summarizes true positives, false positives, and false negatives for the V2 YOLOv8s-seg segmentation model at the selected operating point. The TP, FP, and FN values were computed by matching model predictions against the manual deep-quilting polygon annotations. Since one image can contain true positives, false positives, and false negatives at the same time, Table 8 reports both instance-level counts and image-level counts.

For quality control interpretation, false negatives are more critical than false positives because a missed deep-quilting structure may leave affected material unflagged. False positives mainly increase the amount of manual review. For this reason, the table explicitly reports both the number of false-negative instances and the number of images containing at least one false negative.

Fold 1 produced 40 true positives, 40 false positives, and 11 false negatives, corresponding to a precision of 0.500, recall of 0.784, and F2 of 0.704. Fold 2 produced 62 true positives, 44 false

positives, and 42 false negatives, corresponding to a precision of 0.585, recall of 0.596, and F2 of 0.594. Fold 3 produced 71 true positives, 70 false positives, and 45 false negatives, corresponding to a precision of 0.504, recall of 0.612, and F2 of 0.587.

Table 8: TP, FP, and FN counts for the V2 YOLOv8s-seg segmentation model at the selected operating point. False negatives are reported both as instance counts and as the number of images containing at least one false negative.

Fold	Test roll	Images	TP	FP	FN	Precision	Recall	F1	F2	Images with FN
Fold 1	Roll1	400	40	40	11	0.500	0.784	0.611	0.704	11
Fold 2	Roll2	400	62	44	42	0.585	0.596	0.590	0.594	41
Fold 3	Roll3	350	71	70	45	0.504	0.612	0.553	0.587	38

Some false positives were caused by duplicate predictions around the same physical deep quilting structure rather than by completely unrelated detections. Figure 9 shows two examples in which one prediction overlaps the ground-truth structure and is counted as a true positive, while an additional overlapping prediction around the same target region is counted as a false positive. This indicates that object-level precision can be reduced by duplicate detections even when the predicted regions visually correspond to the same defect structure.

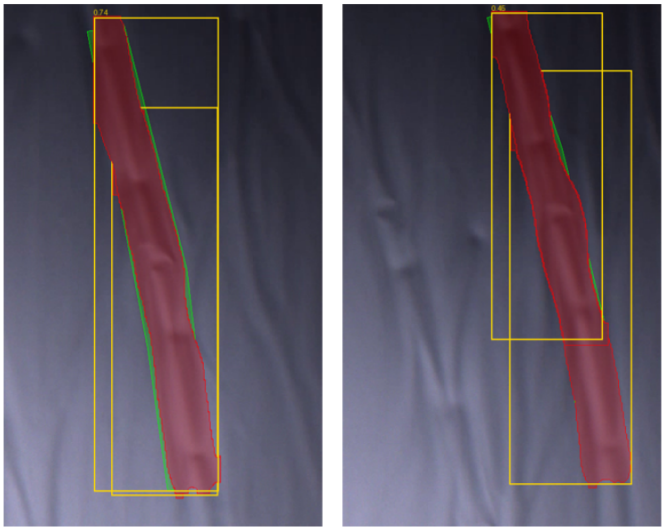


Figure 9: Examples of duplicate predictions around the same deep quilting structure. Green masks indicate ground truth, red masks indicate predicted masks, and yellow boxes indicate predicted bounding boxes. In both examples, one prediction matches the target structure while an additional overlapping prediction is counted as a false positive.

Figure 10 visualizes the TP, FP, and FN counts across the three folds. Fold 1 had the fewest missed detections, while Fold 2 and Fold 3 had higher false-negative counts. These results provide the basis for the later discussion of fold-level difficulty and error patterns, but the causes of the errors are discussed in the Discussion chapter rather than in this Results section.

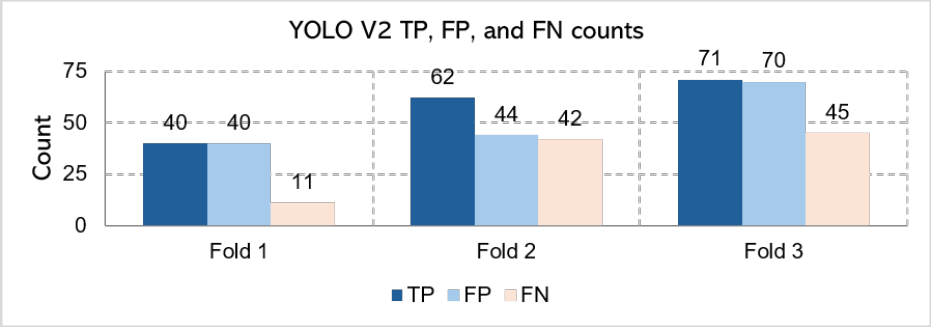


Figure 10: Instance-level true-positive, false-positive, and false-negative counts for V2 YOLOv8s-seg at the selected operating point.

4.6 V3 Node-Aware Verification Results

This section reports the results of the V3 node-aware verification experiment introduced in Section 3.6. As described there, V3 was designed to test whether node evidence could help verify V2 segmentation candidates at the object level. The results are reported in two parts: node detector performance and object-level candidate-handling strategy comparison.

Table 9 shows the node detector results across the three folds. For each fold, the reported epoch was selected by the highest value of `metrics/mAP50-95(B)`. Fold 3 produced the strongest node detector result, with a precision of 0.610, recall of 0.890, mAP50 of 0.698, and mAP50-95 of 0.283. Fold 1 also achieved high recall, 0.883, but with a precision of 0.526. Fold 2 achieved a recall of 0.868 but had the lowest precision and mAP values. Overall, the node detector maintained high recall across folds, while precision and mAP50-95 remained relatively limited. The node detector was therefore more suitable as a structural evidence extractor than as an independent final defect detector.

Table 9: V3 node detector results across the three folds. The reported epoch is selected by the highest box mAP50-95 in the training results.

Fold	Epoch	Precision	Recall	mAP50	mAP50-95
Fold 1	73	0.526	0.883	0.632	0.258
Fold 2	69	0.419	0.868	0.528	0.218
Fold 3	62	0.610	0.890	0.698	0.283

Table 10 reports the object-level strategy comparison. Three candidate-handling strategies were compared. The first strategy, *V2 accept-all candidates*, kept all V2 segmentation candidates and served as the recall-oriented baseline. The second strategy, *node-feature logistic verifier*, used a logistic regression model based on segmentation and node features to accept or reject V2 candidates. The third strategy, *review-oriented node-assisted policy*, assigned candidates to accepted, review-needed, or rejected groups. For evaluation, both accepted and review-needed candidates were counted as surfaced system outputs, because they would still be presented for human review in a quality control workflow.

In Fold 1, the *node-feature logistic verifier* increased precision from 0.360 to 0.467 and increased F2 from 0.684 to 0.726, while recall decreased from 0.882 to 0.843. In Fold 2, the *review-oriented*

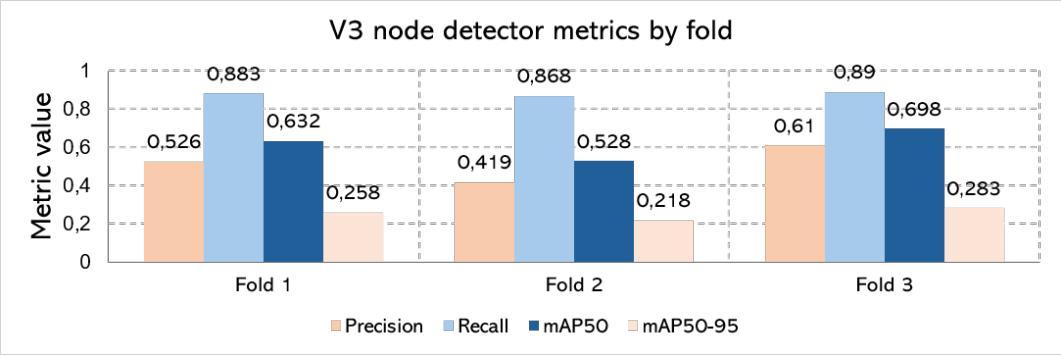


Figure 11: V3 node detector performance across the three folds. The grouped bars show precision, recall, mAP50, and mAP50-95 for node detection.

node-assisted policy slightly improved precision and F2 compared with the *V2 accept-all candidates* baseline, increasing precision from 0.436 to 0.441 and F2 from 0.638 to 0.640. In Fold 3, the surfaced metrics of the *review-oriented node-assisted policy* were the same as the *V2 accept-all candidates* baseline, while the *node-feature logistic verifier* reduced recall and F2.

Table 10: Object-level comparison between the V2 accept-all candidate baseline and the two V3 node-aware candidate-handling strategies.

Fold	Strategy	TP	FP	FN	Precision	Recall	F1	F2
Fold 1	V2 accept-all candidates	45	80	6	0.360	0.882	0.511	0.684
Fold 1	node-feature logistic verifier	43	49	8	0.467	0.843	0.601	0.726
Fold 1	review-oriented node-assisted policy	45	80	6	0.360	0.882	0.511	0.684
Fold 2	V2 accept-all candidates	75	97	29	0.436	0.721	0.543	0.638
Fold 2	node-feature logistic verifier	64	64	40	0.500	0.615	0.552	0.588
Fold 2	review-oriented node-assisted policy	75	95	29	0.441	0.721	0.547	0.640
Fold 3	V2 accept-all candidates	92	177	24	0.342	0.793	0.478	0.628
Fold 3	node-feature logistic verifier	85	152	31	0.359	0.733	0.482	0.606
Fold 3	review-oriented node-assisted policy	92	177	24	0.342	0.793	0.478	0.628

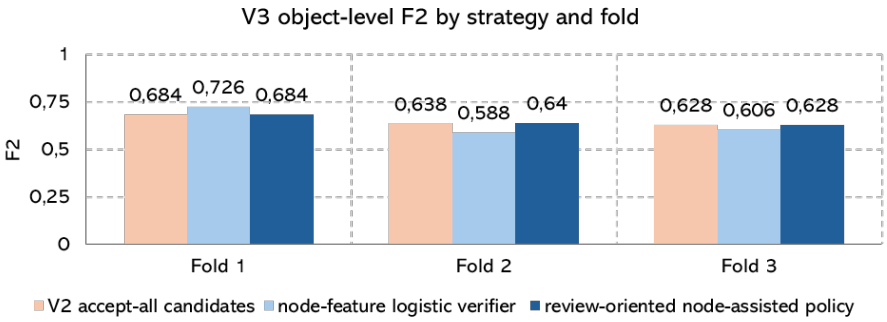


Figure 12: Object-level F2 comparison of the three V3 candidate-handling strategies: V2 accept-all candidates, node-feature logistic verifier, and review-oriented node-assisted policy. Higher F2 indicates better recall-weighted object-level performance.

Overall, the V3 node-aware strategies did not consistently improve object-level precision while preserving recall across all three folds. The final decision was therefore to keep the V2 YOLOv8s-seg segmentation model as the core detection model for roll-level reporting. The *V2 accept-all candidates* strategy was used only as a recall-oriented candidate baseline for the V3 comparison, not as a separate final model. The node-aware strategies provided useful structural evidence for ranking and review support, but they did not yet form a stable automatic filtering system that clearly outperformed the *V2 accept-all candidates* baseline.

4.7 Roll-Level Localization and Reporting Results

This section shows how frame-level predictions were converted into a roll-level quality control report. The report was generated for the selected roll using predicted deep quilting events and summarized them by meter-level event count and density percentage. Event count indicates the number of intersecting events in a meter bin, while density percentage measures the predicted mask union area relative to the inspected crop area in that meter bin.

Table 11 lists the five meter intervals with the highest predicted density. The highest density occurred at 18-19 m, with a density of 6.743% and 8 intersecting events. The second highest interval was 14-15 m, with a density of 6.392%. This interval also had the highest event count, with 12 events. The 28-29 m interval had a density of 6.004% and 9 events. The 10-11 m and 6-7 m intervals also appeared among the highest-density intervals.

Table 11: Highest-density meter intervals in the roll-level report.

Meter index	Start (m)	End (m)	Event count	Density (%)
18	18.0	19.0	8	6.743
14	14.0	15.0	12	6.392
28	28.0	29.0	9	6.004
10	10.0	11.0	6	4.261
6	6.0	7.0	9	4.075

Figure 13 shows the deep quilting event count per meter, and Figure 14 shows the density percentage per meter. Both profiles show that deep quilting was not uniformly distributed along the roll. Instead, several meter intervals concentrated more predicted events or larger predicted mask area. In particular, the 14-15 m, 18-19 m, and 28-29 m intervals were prominent in either count or density and can therefore be prioritized for roll-level quality control review.

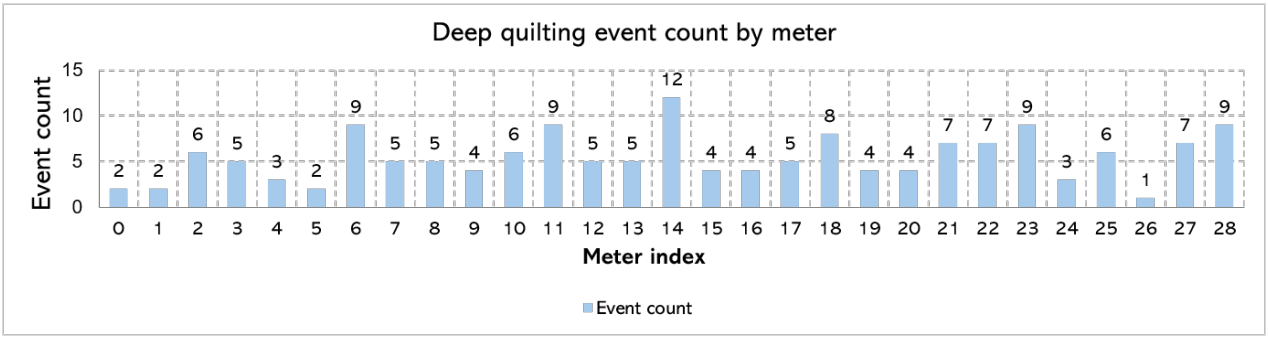


Figure 13: Deep quilting event count per meter for the selected roll. Each bar reports the number of roll-level events intersecting the corresponding meter interval.

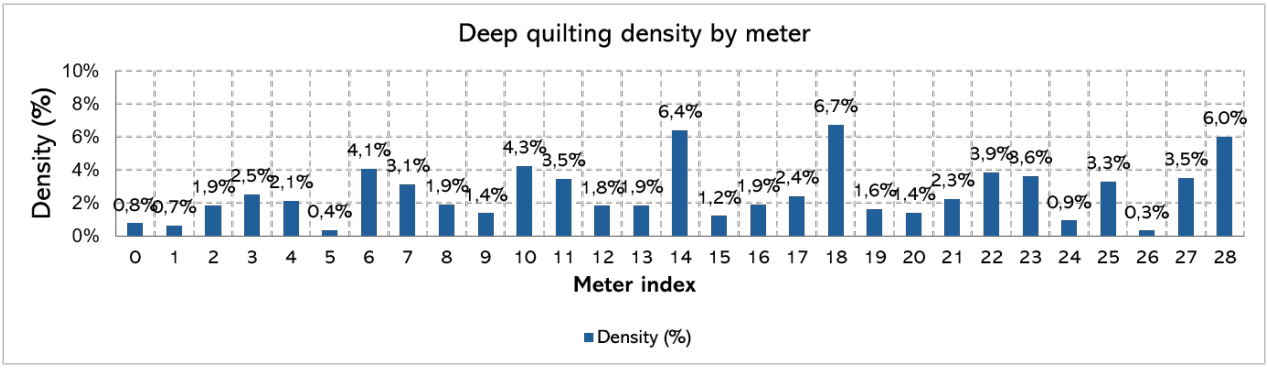


Figure 14: Deep quilting density percentage per meter for the selected roll. Density is computed as the predicted mask union area divided by the inspected crop area within each meter interval.

4.8 Summary of Key Results

The results addressed the main experimental questions of the thesis. First, the dataset construction produced separate data versions for baseline detection, V2 segmentation, and V3 node-aware analysis. Second, the bounding-box baseline learned the deep quilting signal, but its localization quality and recall were limited. Third, V2 YOLOv8s-seg provided the main mask-level detection results and was the most stable model in the comparison with Mask R-CNN and RTMDet-Ins. Fourth, follow-up YOLO tuning did not produce a setting that improved precision, recall, and mask-localization quality at the same time. Fifth, V3 node-aware verification showed that node information had value as structural evidence, but it did not consistently improve final object-level precision while preserving recall. Finally, the roll-level reporting step converted frame-level predictions into per-meter event count and density outputs, making it possible to identify high-risk intervals along the roll.

5 Discussion and Conclusion

This chapter interprets the results, answers the research questions, discusses practical implications and limitations, outlines future work, and concludes the thesis by summarizing the final answer and main contributions.

5.1 Main Findings

The main finding of this thesis is that an automatic computer vision method can detect and localize production-relevant deep quilting from rewinding videos and convert the detections into roll-level quality control statistics. Within the current dataset and imaging setup, the pipeline can separate localized deep-quilting structures from normal quilting background well enough to support industrial review and prioritization. However, the results should not be interpreted as a complete standalone pass/fail method or as a general classifier of all aesthetic defects versus yield-loss defects. The V1 YOLOv8s bounding-box baseline already learned a meaningful deep quilting signal, but its mean mAP50-95 across the three folds was 0.286. This indicates that bounding boxes are limited for representing an elongated and irregular surface structure. In contrast, the V2 YOLOv8s-seg segmentation model achieved mean mask-level precision, recall, mAP50, and mAP50-95 values of 0.673, 0.660, 0.696, and 0.481, respectively. V2 segmentation was, therefore, the strongest modelling result in this thesis.

A second finding is that model choice affected performance under the current data and training setup. The comparison used the primary, untuned three-fold results for the V2 YOLOv8s-seg model, Mask R-CNN, and RTMDet-Ins. Under this setup, YOLOv8s-seg was the most stable of the tested segmentation approaches. Mask R-CNN achieved high recall in some folds but produced many false positives in Fold 1 and an extremely large number of false-positive predictions in Fold 3. RTMDet-Ins did not produce useful true positive detections under the current setup. These results should therefore be interpreted as a practical comparison under the present experimental conditions, rather than as an exhaustive benchmark of all possible instance segmentation architectures.

A third finding is that V3 node-aware verification provided useful structural evidence but did not yet produce a consistently stronger final decision system. The node detector learned the bamboo-like node pattern and maintained high recall across folds. However, the object-level comparison showed that the *node-feature logistic verifier* and the *review-oriented node-assisted policy* did not consistently improve precision while preserving recall. The review-oriented policy was preferable to strict automatic filtering because it avoided automatically discarding potentially relevant defects. However, its surfaced-output performance remained close to the *V2 accept-all candidates* baseline. For this reason, the final system keeps V2 YOLOv8s-seg segmentation as the core detection model, while node-aware verification is treated as exploratory support for candidate ranking and review.

Finally, the roll-level reporting step was the main practical contribution of the pipeline. Frame-level predictions were merged across nearby frames and mapped to material-length coordinates. This produced meter-level event counts and density estimates, which allowed high-risk meter intervals such as 18-19 m, 14-15 m, and 28-29 m to be identified. These outputs show that individual image predictions can be transformed into roll-level statistics that are closer to the needs of quality control review.

5.2 Answering the Research Questions

The main research question asked whether a computer vision pipeline can detect and localize deep quilting or tension wrinkles along the full roll length using rewinding videos and ground-truth annotations. The results support a qualified yes within the scope of the current dataset and imaging setup. The pipeline extracted frames from rewinding videos, cropped a predefined central inspection area, trained segmentation models, generated frame-level predictions, merged predictions across nearby frames, and produced roll-level quality control statistics. The first sub-question concerned whether the data preparation and evaluation design could support reliable model assessment. The results show that the roll-based split was necessary and feasible. Consecutive video frames are visually similar, so a random frame-level split would risk evaluation leakage. The three main rolls were therefore rotated as held-out test rolls, while the practice roll was used only for training augmentation. Although the number of rolls was limited, this design was more appropriate for cross-roll evaluation than a random frame-level split. The second sub-question concerned whether deep quilting could be detected visually. Both the bounding-box baseline and V2 segmentation model showed that the target signal was learnable from cropped inspection images. However, the bounding-box baseline had limited recall and strict localization quality. V2 segmentation produced valid mask-level results across all three folds and achieved higher mAP50-95 than the baseline. Deep quilting can therefore be detected, but it is better represented by segmentation masks than by bounding boxes. The third sub-question concerned whether node information could improve the detection pipeline. The current answer is limited. V3 showed that a node detector can learn the node pattern, and that node information can provide structural evidence for candidate ranking or triage. However, V3 did not consistently improve object-level precision while preserving recall. Node-aware verification is therefore better interpreted as a review-support layer than as a final automatic rejection layer. The fourth sub-question concerned whether frame-level predictions could be converted into a roll-level quality control report. The results give a positive answer. Meter-level event count and density profiles identified intervals where predicted deep quilting was more concentrated. This output is more useful for quality control than isolated frame-level predictions, because it indicates where defects occur along the roll and where review should be prioritized.

5.3 Relation to Earlier Research

The findings are consistent with the broader literature on automated visual inspection, which treats industrial inspection as a system-level problem involving acquisition, defect definition, detection method, and deployment context [NJ95, MPZ+03]. The present results support this view. The most useful output was not a single model score, but an inspection pipeline combining central crop selection, roll-based evaluation, prediction merging, and roll-level reporting. This is especially relevant for continuous materials, where motion, lighting, surface texture, and production constraints affect inspection design [NMD14, TCSL23]. The results also support previous work on segmentation-based surface defect detection. Segmentation is particularly useful when defect shape, area, and spatial distribution are important [TŠSS20]. Deep quilting is elongated, irregular, and locally variable, so a bounding box cannot represent its extent accurately. The improvement of V2 over the bounding-box baseline is consistent with this interpretation and with surface-defect studies that emphasize localization under complex texture conditions [LDJ+20]. At the same time, the V3 results provide a more cautious view of high-recall candidate generation. In industrial

quality control, missing defects can be more problematic than producing additional candidates for review, which makes a high-recall candidate set attractive [RPZ+22]. However, the present results show that high recall alone does not guarantee that a node-aware verifier will improve final precision. In the current setting, node information is better used for ranking and triage than for hard filtering. The roll-level reporting step is also related to the general principle of combining overlapping visual observations into a consistent output. Classical image alignment and stitching methods use overlapping observations to build a coherent result [Sze07]. This thesis does not perform panoramic stitching, but it applies a related idea by merging nearby frame-level predictions into roll-level events. The resulting output is not a stitched image but an event list, event-count profile, and density profile along the roll.

5.4 Practical Implications for Quality Control

From a practical perspective, the value of the system is not only that it detects deep quilting in individual cropped inspection images but also that it converts visual predictions into roll-level quality control information. In continuous material production, quality control users need to know where defects occur along the roll, how frequently they occur, and which intervals should be inspected first. Meter-level event count and density profiles directly support this need. The V2 segmentation output can be used to prioritize high-risk intervals. For example, the roll-level report identified 18-19 m as the highest-density interval and 14-15 m as the interval with the highest event count. These intervals provide concrete locations for review and reduce the need for frame-by-frame manual inspection. The report therefore improves interpretability by linking model predictions to material position. However, the current system should be interpreted as decision support rather than as an automatic rejection system. First, the segmentation model still produces false positives and false negatives, and false negatives are especially important because missed defects may leave affected material unflagged. Second, V3 node-aware verification did not yet produce a stable improvement in final object-level decisions. Third, the data were collected from a limited number of rolls. For these reasons, the current output is most appropriate for review prioritization, reporting, and inspection support rather than as a standalone pass/fail criterion.

To turn this output into a pass/fail method, additional industrial validation would be required. In particular, roll-level event counts, density profiles, and high-risk intervals should be compared with downstream processing or yield data. This would make it possible to determine whether specific deep-quilting levels are associated with unacceptable material performance and to define validated pass/fail specifications. Until such specifications are established, the most appropriate use of the system is to support inspection decisions and prioritize review rather than to automatically reject material.

5.5 Interpretation of the V2 Segmentation Results

The success of V2 segmentation can be explained by the visual structure of deep quilting. Deep quilting is not a compact rectangular object. It is an elongated, irregular, and locally variable surface defect. A mask can represent the defect region more directly than a bounding box and can be used for density calculation. The comparison between the baseline and V2 supports this explanation, because V2 achieved higher mAP50-95 values than the bounding-box baseline in all three folds. The fold-level variation also shows that the task still has generalization difficulty. Fold

1 and Fold 3 performed more strongly than Fold 2. This may reflect differences in the held-out roll appearance, defect morphology, annotation distribution, or background texture. The important point is that roll-based evaluation exposed these differences. This is useful because industrial inspection systems have to generalize to unseen rolls rather than only to near-duplicate frames. The follow-up YOLO tuning results further indicate that hyperparameter adjustment alone did not solve the trade-off between precision, recall, and mask-localization quality. The best mAP50-95 tuned run only slightly improved mAP50-95 but reduced recall. The recall-oriented run improved recall but reduced mAP50-95. Future improvements are therefore more likely to come from more roll diversity, more consistent annotation, and better data coverage than from small tuning changes alone.

5.6 Interpretation of the V3 Node-Aware Results

The V3 results should be interpreted as exploratory rather than final. The node detector achieved high recall across folds, indicating that the bamboo-like node pattern is learnable. This supports the visual assumption that deep quilting contains useful nodal structure, and it shows that the node labels were not merely noise. However, node information did not consistently translate into better final object-level decisions. The node-feature logistic verifier strategy improved precision and F2 in Fold 1, but it did not show a stable advantage in Fold 2 and Fold 3. The review-oriented node-assisted policy results were mostly close to the V2 accept-all candidates baseline. This suggests that the current node features are useful for ranking but insufficient for reliable hard acceptance or rejection. This interpretation is important for quality control. Since missed defects are more problematic than additional review candidates, hard filtering is risky. An accept/review/reject triage structure fits the current evidence better. High-confidence candidates can be surfaced with priority, uncertain candidates can remain available for human review, and only clearly weak candidates should be deprioritized. In this form, V3 supports the V2 segmentation pipeline rather than replacing it.

5.7 Limitations and Threats to Validity

This study has several limitations. First, the number of rolls was limited. Although the evaluation used a roll-based three-fold split, the three main rolls cannot represent all future production variability. The results should therefore not be overstated as evidence of stable generalization to all future rolls. A related limitation is transferability. The pipeline was developed for deep quilting in the current silicon anode material and imaging setup. It should not be assumed to detect other defect types without additional annotation, training, and validation. Other defects may have different visual appearances, spatial scales, severity definitions, and quality control relevance. Similarly, changes in product generation, material width, camera field of view, or lighting conditions would require recalibration of the central inspection crop, pixel-to-millimetre conversion, and roll-position mapping before the system could be reused reliably. Second, the annotation protocol affects the validity of the results. The boundary between normal quilting and deep quilting is sometimes visually ambiguous, especially for weak, partial, or borderline structures. The operational definition improved consistency, but no formal inter-annotator agreement was reported. Annotation uncertainty therefore remains a threat to internal validity. Third, the node labels used in V3 were auxiliary structural labels, not exhaustive annotations of all visible nodes. The

node detector should therefore be interpreted as a detector of deep-quilting-related node evidence, not as a complete general node detector. This limits the conclusions that can be drawn from V3. Fourth, the roll-level localization used an approximately constant-speed assumption. This was reasonable for the present setting because the material speed during rewinding was considered stable. However, speed fluctuation, slippage, or encoder mismatch could introduce position errors in future deployments. The roll-level position estimates should therefore be validated under additional production conditions. Finally, the evaluation levels must not be mixed. V2 region-level metrics, V3 object-level metrics, and roll-level density statistics answer different questions. Treating them as one performance number would lead to an incorrect interpretation. This thesis therefore separated model-level, object-level, and roll-level evaluation.

5.8 Future Work

Future work should follow two connected directions: research validation and industrial deployment. First, the roll-level dataset should be expanded. The current dataset is still limited in the number of rolls, normal examples, borderline cases, and imaging conditions it covers. More data should therefore be accumulated over time to evaluate cross-roll generalization more reliably. This is especially important for distinguishing normal quilting backgrounds from true deep quilting under variable production conditions. The annotation protocol should also be strengthened through double annotation, inter-annotator agreement analysis, and uncertainty labels. Weak or partial defects could be annotated with certainty attributes rather than forced into a binary positive or negative label.

Second, the current inspection output should be linked to downstream production data. Roll-level event counts, density profiles, and high-risk intervals should be compared with downstream processing or yield outcomes. This step is necessary before defining pass/fail specifications, because the present thesis evaluates visual detection and localization but does not establish which defect levels cause unacceptable material performance. Once such correlations are available, validated threshold rules could be developed for pass/fail or review-needed decisions.

Some next steps can be carried out as engineering or technician work once the workflow is fixed. These include recording additional rolls under known settings, running the existing frame extraction and prediction pipeline, checking the generated roll-level reports, and collecting human review feedback for high-risk intervals. Additional research work would be needed for new defect categories, yield-prediction modelling, major changes to the model architecture, or validation under substantially different product and imaging conditions.

The scope of the system could also be extended to other defect types, but this would require new defect definitions, representative annotations, and separate validation. The present model was trained for deep quilting and should not be interpreted as a general defect detector. Other defects may require different labels, thresholds, model settings, and evaluation criteria. Future product or imaging changes should also be treated as new validation conditions. Wider rolls, different material-strip geometry, changed camera position, or changed lighting could affect the central inspection crop, pixel-to-millimetre conversion, and roll-position mapping. For such cases, the inspection pipeline would need recalibration and revalidation before being used for quality control decisions.

Finally, the V3 node-aware layer could be improved with more structured node features. Future features could include node spacing, node alignment, node span relative to the segmentation mask,

and sequence continuity. These features may better capture the physical structure of deep quilting than simple node count or confidence summaries. The system could also be developed into a human-in-the-loop quality control interface showing high-risk meter intervals, representative frames, candidate confidence, node evidence, and density profiles. Human feedback from this interface could then be reused for annotation refinement, model monitoring, and periodic retraining when product ranges or production conditions change.

5.9 Final Answer to the Research Question

Within the scope of the current dataset and imaging setup, rewinding videos and ground-truth annotations can be used to build a structured inspection pipeline for detecting and localizing deep quilting along a full roll. The pipeline extracts cropped inspection images from rewinding videos, trains segmentation models, generates frame-level deep quilting predictions, merges predictions across nearby frames, and reports meter-level event count and density.

The strongest modelling result is the V2 YOLO segmentation model. It outperforms the earlier bounding-box baseline in strict localization quality and provides masks that are more suitable for density-based roll-level reporting. V3 node-aware verification provides useful structural evidence, but it does not yet consistently improve object-level precision while preserving recall. Therefore, the final practical system should be interpreted as segmentation-based detection with roll-level reporting, supported by exploratory node-aware triage rather than fully automatic node-based rejection.

5.10 Main Contributions

This thesis contributes a structured inspection pipeline that connects video-based surface inspection with roll-level quality control reporting. The main methodological contributions are the roll-based evaluation design, the comparison between bounding-box detection, segmentation, and node-aware verification, and the conversion of frame-level predictions into meter-level event count and density statistics.

The main contribution of this thesis is therefore not only a trained deep quilting detector but also a structured inspection pipeline that links visual defect detection to roll-level quality control reporting. In the current setting, segmentation is the most effective core model, while node-aware verification is better treated as supporting evidence for review and triage.

References

[Bra00] Gary Bradski. The opencv library. *Dr. Dobb's Journal of Software Tools*, 2000.

[CVA23] CVAT.ai Corporation. Computer Vision Annotation Tool (CVAT), 2023. Computer software.

[CWP+19] Kai Chen, Jiaqi Wang, Jiangmiao Pang, Yuhang Cao, Yu Xiong, Xiaoxiao Li, Shuyang Sun, Wansen Feng, Ziwei Liu, Jiarui Xu, Zheng Zhang, Dazhi Cheng, Chenchen Zhu, Tianheng Cheng, Qijie Zhao, Buyu Li, Xin Lu, Rui Zhu, Yue Wu, Jifeng Dai, Jingdong Wang, Jianping Shi, Wanli Ouyang, Chen Change Loy, and Dahua Lin. MMDetection: Open mmlab detection toolbox and benchmark. In *arXiv preprint arXiv:1906.07155*, 2019.

[FR10] Per-Erik Forssén and Erik Ringaby. Rectifying rolling shutter video from hand-held devices. In *2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pages 507–514. IEEE, 2010.

[HGDG17] Kaiming He, Georgia Gkioxari, Piotr Dollár, and Ross Girshick. Mask r-cnn. In *Proceedings of the IEEE international conference on computer vision*, pages 2961–2969, 2017.

[JCQ23] Glenn Jocher, Ayush Chaurasia, and Jing Qiu. Ultralytics YOLOv8. Computer software, 2023.

[LCC08] Chia-Kai Liang, Li-Wen Chang, and Homer H Chen. Analysis and compensation of rolling shutter effect. *IEEE transactions on image processing*, 17(8):1323–1330, 2008.

[LDJ+20] Xiaoming Lv, Fajie Duan, Jia-jia Jiang, Xiao Fu, and Lin Gan. Deep metallic surface defect detection: The new benchmark and detection network. *Sensors*, 20(6):1562, 2020.

[LSD15] Jonathan Long, Evan Shelhamer, and Trevor Darrell. Fully convolutional networks for semantic segmentation. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 3431–3440, 2015.

[LZH+22] Chengqi Lyu, Wenwei Zhang, Haian Huang, Yue Zhou, Yudong Wang, Yanyi Liu, Shilong Zhang, and Kai Chen. RTMDet: An Empirical Study of Designing Real-Time Object Detectors. arXiv preprint, 2022.

[MPZ+03] Elias N Malamas, Euripides GM Petrakis, Michalis Zervakis, Laurent Petit, and Jean-Didier Legat. A survey on industrial vision systems, applications and tools. *Image and vision computing*, 21(2):171–188, 2003.

[NJ95] Timothy S Newman and Anil K Jain. A survey of automated visual inspection. *Computer vision and image understanding*, 61(2):231–262, 1995.

[NMD14] Nirbhar Neogi, Dusmanta K Mohanta, and Pranab K Dutta. Review of vision-based steel surface inspection systems. *EURASIP Journal on Image and Video Processing*, 2014(1):50, 2014.

[PGM+19] Adam Paszke, Sam Gross, Francisco Massa, Adam Lerer, James Bradbury, Gregory Chanan, Trevor Killeen, Zeming Lin, Natalia Gimelshein, Luca Antiga, Alban Desmaison, Andreas Kopf, Edward Yang, Zachary DeVito, Martin Raison, Alykhan Tejani, Sasank Chilamkurthy, Benoit Steiner, Lu Fang, Junjie Bai, and Soumith Chintala. Pytorch: An imperative style, high-performance deep learning library. In *Advances in Neural Information Processing Systems*, volume 32, 2019.

[PVG+11] Fabian Pedregosa, Gaël Varoquaux, Alexandre Gramfort, Vincent Michel, Bertrand Thirion, Olivier Grisel, Mathieu Blondel, Peter Prettenhofer, Ron Weiss, Vincent Dubourg, et al. Scikit-learn: Machine learning in python. *the Journal of machine Learning research*, 12:2825–2830, 2011.

[RBC+17] David R Roberts, Volker Bahn, Simone Ciuti, Mark S Boyce, Jane Elith, Gurutzeta Guillera-Arroita, Severin Hauenstein, José J Lahoz-Monfort, Boris Schröder, Wilfried Thuiller, et al. Cross-validation strategies for data with temporal, spatial, hierarchical, or phylogenetic structure. *Ecography*, 40(8):913–929, 2017.

[RHGS15] Shaoqing Ren, Kaiming He, Ross Girshick, and Jian Sun. Faster r-cnn: Towards real-time object detection with region proposal networks. *Advances in neural information processing systems*, 28, 2015.

[RPZ+22] Karsten Roth, Latha Pemula, Joaquin Zepeda, Bernhard Schölkopf, Thomas Brox, and Peter Gehler. Towards total recall in industrial anomaly detection. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 14318–14328, 2022.

[Sze07] Richard Szeliski. Image alignment and stitching: A tutorial. *Foundations and trends in computer graphics and vision*, 2(1):1–104, 2007.

[TCSL23] Bo Tang, Li Chen, Wei Sun, and Zhong-kang Lin. Review of surface defect detection of steel products based on machine vision. *IET image processing*, 17(2):303–322, 2023.

[TŠSS20] Domen Tabernik, Samo Šela, Jure Skvarč, and Danijel Skočaj. Segmentation-based deep-learning approach for surface-defect detection. *Journal of Intelligent Manufacturing*, 31(3):759–776, 2020.