INSTITUTO
UNIVERSITÁRIO
DE LISBOA

A Vision-Based System to Guide the Drilling Process in Cork Stopper Production

Gonçalo Miguel Inácio do Amaral Botelho

Master in Computer Engineering

Supervisor:
PhD Pedro Figueiredo Santana, Associate Professor,
Iscte – Instituto Universitário de Lisboa

Supervisor:
PhD Tomás Gomes Silva Serpa Brandão, Assistant Professor,
Iscte – Instituto Universitário de Lisboa

October, 2025

TECHNOLOGY
AND ARCHITECTURE

Department of Information Science and Technology

A Vision-Based System to Guide the Drilling Process in Cork Stopper Production

Gonçalo Miguel Inácio do Amaral Botelho

Master in Computer Engineering

Supervisor:
PhD Pedro Figueiredo Santana, Associate Professor,
Iscte – Instituto Universitário de Lisboa

Supervisor:
PhD Tomás Gomes Silva Serpa Brandão, Assistant Professor,
Iscte – Instituto Universitário de Lisboa

October, 2025

To my family, friends and girlfriend,
for their constant support, understanding and motivation throughout this journey

Acknowledgment

I would like to express my deepest gratitude to my supervisors, Dr. Pedro Santana and Dr. Tomás Brandão, for their constant support throughout the development of this dissertation. Their availability, guidance and encouragement were essential in helping me reach the final outcome of this work.

I am also grateful to Introsys, S.A. for providing the dataset, which was a valuable asset to this research. In particular, I would like to thank Engineer Magno Guedes for his availability and assistance.

To my family, I owe my sincere appreciation for their unconditional love, patience and support. Without them, this dissertation would not have been possible.

I also wish to thank my friends, especially Hugo and Bruno, for their understanding and encouragement throughout this challenging journey.

To my colleague and friend João, thank you for accompanying and motivating me during the entire process.

Finally, my deepest thanks go to my girlfriend Beatriz, who has been my pillar through every moment, even the most difficult ones and the main reason this dissertation has reached completion.

Resumo

Esta dissertação aborda a automatização do processo de furação de tiras de cortiça para a produção de rolhas, através da aplicação de técnicas de visão computacional. O trabalho propõe um sistema capaz de reproduzir os padrões de furação tradicionalmente realizados por operadores humanos, com o objetivo de aumentar a eficiência e reduzir riscos laborais e desperdícios de matéria-prima. Foi adotada a metodologia *Design Science Research Methodology* (DSRM), que orientou a identificação do problema, o desenvolvimento do artefacto e a sua avaliação. O sistema desenvolvido baseia-se numa abordagem de *deep learning* supervisionado, envolvendo deteção, segmentação e projeção de furos em imagens de tiras não furadas, alinhamento entre superfícies e filtragem dos resultados. Os testes efetuados com imagens provenientes de um contexto real de produção demonstraram que o modelo atinge um desempenho elevado, com *Precision* de 0.861, *Recall* de 0.911 e *F1-Score* de 0.8855, evidenciando a capacidade do sistema em reproduzir com elevada fidelidade os padrões de furação humanos. Estes resultados validam o potencial da visão computacional na automatização de tarefas industriais de alta precisão.

Palavras Chave: *visão computacional* , *cortiça*, *automatização industrial*, *deteção de objetos*, *segmentação de imagens*, *aprendizagem profunda*

Abstract

This dissertation addresses the automation of the drilling process of cork strips for cork stopper production through the application of computer vision techniques. The work proposes a system capable of reproducing the drilling patterns traditionally performed by human operators, aiming to increase efficiency and reduce occupational risks and material waste. The *Design Science Research Methodology* (DSRM) was adopted to guide the problem identification, artifact development and evaluation stages. The developed system is based on a supervised *deep learning* approach, involving the detection, segmentation and projection of holes in images of undrilled cork strips, alignment between surfaces and filtering of the results. Tests conducted with images collected in a real production context demonstrated that the model achieved high performance, with *Precision* of 0.861, *Recall* of 0.911 and *F1-Score* of 0.8855, highlighting the system's ability to accurately reproduce human drilling patterns. These results validate the potential of computer vision in automating high-precision industrial tasks.

KEYWORDS: *computer vision*, *cork*, *industrial automation*, *object detection*, *image segmentation*, *deep learning*

Contents

Acknowledgment iii

Resumo v

Abstract vii

List of Tables xi

List of Figures xiii

List of Acronyms xv

Chapter 1. Introduction 1
1.1. Background Motivation 1
1.2. Research Questions 4
1.3. Goals 4
1.4. Methodology 5
1.5. Dissertation Outline 7

Chapter 2. Literature Review 9
2.1. Background Concepts 11
2.1.1. Defects Characterization 11
2.1.2. Cork Stoppers Punching 12
2.1.3. Computer Vision, Object Detection and Image Segmentation 13
2.2. Related Work 13
2.3. Contribution 22

Chapter 3. System Design and Development 25
3.1. System Overview 25
3.2. Dataset Analysis 27
3.3. System Development Stages 28
3.3.1. Cork Strip Detection and Segmentation 30
3.3.2. Holes Detection and Segmentation 34
3.3.3. Holes Projection onto the undrilled Images 35
3.3.4. Undrilled Strip Alignment 38
3.3.5. Predicted Holes Filtering 39
3.4. Predictive Model Design and Implementation 39
3.5. Software Architecture and Implementation Overview 41

3.6. Hole Prediction Pipeline 42

Chapter 4. Results and Discussion 45
4.1. Evaluation Metrics 45
4.1.1. Segmentation Level Assessment Metrics 46
4.1.2. Geometric Level Assessment Metrics 47
4.2. Segmentation Level Evaluation 48
4.3. Geometric Level Evaluation 50
4.4. Qualitative Results 52
4.5. Discussion 54

Chapter 5. Conclusions and Future Work 57
5.1. Conclusions 57
5.2. Future Work 59

References 61

Appendix A. Software Architecture Diagram 65

Appendix B. End-to-End Application Example on a Cork Strip 67
B.1. Original Images of the Cork Strip 67
B.2. Cork Strip Detection and Segmentation 67
B.3. Holes Detection and Segmentation 67
B.4. Holes Projection onto the Undrilled Images 68
B.5. "Good Cork" Regions Prediction 68
B.6. Undrilled Images Alignment 69
B.7. Final Predicted Holes 69

List of Tables

1.1 Dissertation Schedule 8

2.1 Distribution of Publication Year 10
2.2 Distribution of Publication Year *ad hoc* search for RQ1 11

3.1 Summary of the selected dataset 28
3.2 Statistics of the circle radii and distances between centers, in pixels (Pixels (Px)) 35

4.1 Aggregated results on the test set (Hungarian matching; tolerance = 60 px) 50
4.2 Summary of hole counts — intersection (images present in GT and Pred) 51

List of Figures

1.1 Cork stoppers punching (traditional method) [3] 3
1.2 DSRM Process Model [6] 6

2.1 Examples of defects [8] 12
2.2 Examples of belly and back [9] 12

3.1 Left view of the data acquisition setup. The cork strip is positioned on the table between the two cameras. 25
3.2 Representation of the top view of the acquisition setup showing the relative position of the left and right cameras, the table and the strip. 26
3.3 BPMN diagram of the System's workflow 26
3.4 Example of the four views of a cork strip 28
3.5 Overview of the system development stages 29
3.6 Examples of binary masks resulting from the application of background subtraction (Mixture of Gaussians (MOG)) 30
3.7 Examples of intersection between the masks resulting from the improved application of the background subtraction method and the real images 31
3.8 Examples of cork strip detections by the You Only Look Once (YOLO) model 32
3.9 Examples of image segmentation with the YOLO model 33
3.10 BPMN diagram of the strip detection and segmentation Process 33
3.11 Examples of image segmentation with the Segment Anything Model (SAM) model 34
3.12 Examples of hole segmentation 35
3.13 Example of strip images with swapped pairs 36
3.14 Example of strip images with swapped pairs and with deformations 37
3.15 Example of hole projection 38
3.16 Example of alignment of undrilled surface images of the same strip 39
3.17 Classical U-Net architecture diagram [43] 40

4.1 Test set inference sample 49
4.2 Confusion Matrix - Final Evaluation 50
4.3 Distribution of the difference between predicted and real hole counts 51
4.4 Example of hole prediction (1) 52
4.5 Example of hole prediction (2) 52
4.6 Example of hole prediction (3) 53

4.7 Example of hole prediction (4) 53
4.8 Example of hole prediction (5) 53
4.9 Example of hole prediction (6) 54
4.10 Example of hole prediction (7) 54
4.11 Example of hole prediction (8) 54

A.1 Class diagram representing the main modules of the developed system and their interactions. 65

B.1 Original images representing the undrilled and drilled views of the strip 67
B.2 Segmented Images of the Cork Strip 67
B.3 Holes Segmentation Binary Mask (drilled images) 67
B.4 Holes projection onto the undrilled images. Projected hole circumferences marked in red 68
B.5 "Good Cork" regions predicted in both undrilled images. Contours of the “Good Cork” regions marked in red and circumferences of the ground truth holes marked in green 68
B.6 Overlay of undrilled images binary masks after alignment 69
B.7 Overlay of predicted holes (after filtering) and ground truth holes. Ground truth holes are marked in green, true positive predictions in blue and yellow lines indicate the displacement between each matched predicted and ground truth hole 69

List of Acronyms

AI: Artificial Intelligence

ASPP: Atrous Spatial Pyramid Pooling

BCE: Binary Cross Entropy

BiFPN: Bi-Directional Feature Pyramid Network

BPMN: Business Process Model and Notation

CA: Coordinate Attention

CC-ICA: Class-Conditional Independent Component Analysis

CDA: Canonical Discriminant Analysis

CID: Cork Image Database

CNN: Convolutional Neural Network

CV: Computer Vision

DS: Design Science

DSRM: Design Science Research Methodology

FLANN: Fast Library for Approximate Nearest Neighbors

GAN: Generative Adversarial Network

GLCM: Gray-Level Co-Occurrence Matrix

GMG: Geometric Mixture of Gaussians

GPU: Graphics Processing Unit

IC: Inclusion Criteria

INE: Instituto Nacional de Estatística

IOU: Intersection Over Union

IS: Information Systems

LSH: Locality Sensitive Hashing

MOG: Mixture of Gaussians

ORB: Oriented FAST and Rotated BRIEF

Px: Pixels

R-CNN: Regions-based Convolutional Neural Network

SAM: Segment Anything Model

SDA: Stepwise Discriminant Analysis

SLR: Systematic Literature Review

SVM: Support Vector Machine

SVR: Support Vector Regression

UMS-FRCNN: Multi-Scale Faster R-CNN with Upper Lower Layer Fusion

WWF: World Wide Fund for Nature

YOLO: You Only Look Once

ZAC: Zone of Acceptable Cork

ZPC: Zone of Possible Centers

CHAPTER 1

Introduction

1.1. Background Motivation

The cork industry plays a fundamental role in the global economy, standing out in countries of the Western Mediterranean basin. The "*Montados de Sobro*" (forests of cork oaks, the tree from which cork is extracted) occupy an estimated area of more than 2.1 million hectares in these regions, forming one of the most emblematic agrosilvopastoral systems in the world. Every year, the extraction of around 200 thousand tons of cork is estimated. One third of the global area of cork oaks belongs to Portugal (22% of the country's forest area), thus being the world's largest producer of cork [1].

These cork oak landscapes are not only a cornerstone of Portugal's natural heritage but also ecosystems of international relevance. They extend across Spain, France, Italy, Morocco, Algeria and Tunisia, where they provide a balance between environmental preservation and economic activity [2]. Known as *montado*, in Portugal and *dehesa*, in Spain, this systems are recognised by the World Wide Fund for Nature (WWF) and the Food and Agriculture Organization as examples of sustainable land management that combine tree cover, grazing and small scale agriculture while maintaining high levels of biodiversity [2]. They are home to endangered species such as the Iberian lynx and the Iberian imperial eagle and are among the most diverse forest habitats in Europe. Ecologically, they contribute to climate regulation, carbon sequestration, soil protection and water retention, preventing desertification in semi-arid regions of southern Europe and North Africa [2].

One of the most remarkable aspects of cork is the sustainability of its extraction process. The bark of the cork oak can be removed without felling or harming the tree, allowing it to regenerate and be harvested again every 9 to 12 years [3]. This cyclical process ensures that the forest remains intact, acting as a long-term carbon sink while supporting a renewable industry. According to the WWF, cork oak forests are "a rare example of an industrial system that depends on the survival of the ecosystem itself," meaning that forest conservation and economic activity are interdependent [2]. This relationship has safeguarded millions of hectares of Mediterranean woodland that might otherwise have been converted or abandoned.

From a socio-economic standpoint, cork production sustains rural livelihoods across southern Europe and North Africa. It is estimated that thousands of people depend directly or indirectly on cork oak landscapes for their income: from bark stripping to processing, livestock management and the production of finished goods [2]. This makes the cork sector an essential pillar for maintaining rural employment and combating depopulation in areas with limited economic alternatives.

Between 2011 and 2022, global cork exports increased by almost 60%, reaching around 2000 million euros. Portugal is responsible for about 60% of exports, followed by Spain (19.8%) and France (5.2%). Thus, Portuguese exports of cork and its products exceeded 1200 million euros (representing 1.5% of the total of its exports of goods), with a positive balance of 1000 million euros. Although the quantity exported has not increased significantly, this result is supported by the increase in the average value per kilogram. According to Instituto Nacional de Estatística (INE), in 2021, there were 815 companies in Portugal linked to the industry, employing 8555 people. The sector is structured into three main subclasses: cork preparation industry, manufacture of cork stoppers and manufacture of other cork products. Companies linked to the manufacture of cork stoppers account for more than half of the sectoral composition of the industry.

In recent years, cork has achieved greater diversification and international recognition. Beyond its traditional use in wine closures, cork's exceptional properties, such as low density, elasticity, impermeability, thermal and acoustic insulation, vibration damping and fire resistance, have made it a valuable material in construction, transport, aerospace and design [4] [1]. In the building sector, cork is used in insulation panels, flooring and exterior coatings. In aerospace and automotive industries, it serves as a lightweight, high-performance composite for vibration and heat protection. Moreover, in fashion and design, cork fabrics are increasingly promoted as a sustainable alternative to leather and plastics [4]. Despite this diversification, cork stoppers continue to represent around 70% of the industry's total commercial value, confirming their dominant economic importance [2] [4].

Over the last decade, investment in the industry has also increased, reaching more than 80 million euros annually, representing about 2% of investment in manufacturing industries. The manufacture of stoppers accounts for a large part of this amount, reaching an annual investment of around 50 million euros [4]. Its production process involves several stages, from stripping (removal of cork from the cork oaks) to quality control of the stoppers produced. As it is a natural resource, cork may present defects, such as cracks or insect damage. Their presence in cork stoppers undermines their quality, directly impacting their value. Thus, in most cases, its production requires the intervention of highly qualified human labor to ensure that only defect-free areas are used for punching. Currently, as in the industry in general, there is a trend towards evolving the production process, namely with the automation of some of the stages. Even so, with the need to guarantee the quality of the stoppers (particularly the higher-value ones), human intervention remains crucial. One of these stages is the drilling of cylindrical stoppers. This process requires from the operator great physical and mental endurance, as well as good visual and motor coordination. The operator analyses the cork strip, a rectangular section of bark with a width slightly larger than the stopper's intended diameter and manually positions it next to a tubular driller, which is operated by foot [3], as shown in Figure 1.1. The operator must continuously evaluate the strip to avoid areas of visible

defects or irregular density. As a result, the process is repetitive, technically complex and physically exhausting. The operator's hands remain close to the drilling mechanism, creating a constant risk of injury. According to industrial data, an experienced operator can typically produce around 16,000 stoppers during an eight-hour shift [5]. To increase productivity and address safety concerns, automated drilling systems have been developed. These systems use robotic feeding arms to introduce cork strips into automatic drilling machines, achieving production rates of up to 35,000 stoppers per day [5].



FIGURE 1.1. Cork stoppers punching (traditional method) [3]

Even so, with the need to guarantee the quality of the stoppers (particularly the higher-value ones), human intervention remains crucial. Such approaches prioritize throughput and uniformity rather than the qualitative selection of the cork regions to be used [5]. The machine treats the strip as a homogeneous material, disregarding natural imperfections. Consequently, these systems tend to generate more waste and produce stoppers with greater variability in quality compared to the manual approach. For high-value natural cork stoppers, typically used in premium wine markets, manual or semi-manual drilling remains the standard practice because it allows operators to make nuanced decisions that preserve both yield and quality.

The central challenge of the cork sector lies in maintaining the delicate balance between productivity and quality. Achieving this equilibrium is essential not only for the economic competitiveness of cork against synthetic alternatives but also for the preservation of the cork oak forests themselves, which remain economically viable only while cork retains its market value. Therefore, automating the drilling process through an intelligent vision-based system, capable of analysing cork strips and generating drilling patterns that emulate the operator's expertise, has both industrial and environmental significance. Such an approach would optimise material use, reduce workplace risks and enhance production efficiency without compromising quality. At a broader level, innovation in this domain directly contributes to the sustainability of Mediterranean cork ecosystems, the economic resilience of rural communities and the global positioning of cork as a high-value, renewable and environmentally responsible material.

1.2. Research Questions

To provide a clear direction for the investigation, this dissertation is structured around a set of Research Questions (RQ):

(1) RQ1: Can computer vision models effectively reproduce the decision-making process of human operators in cork stopper drilling, achieving drilling patterns comparable to those obtained through manual intervention?
(2) RQ2: How do data preparation and pre-training strategies affect the performance and generalization ability of computer vision models in predicting drilling positions on cork strips?
(3) RQ3: Is it possible to adapt a segmentation model to accurately predict drilling patterns on cork strips, ensuring robustness and consistency across different visual and structural conditions?

These questions explore whether computer vision can effectively reproduce human decision-making in cork stopper drilling, how data preparation and pre-training strategies impact model performance and if a segmentation model is suitable for achieving accurate and robust predictions. Their formulation ensures that the study remains focused on assessing both the feasibility and the technical optimization of the proposed solution.

1.3. Goals

The main goal of this dissertation is to develop and evaluate a computer vision based approach capable of reproducing the manual decision-making process involved in cork stopper drilling, by automatically predicting drilling positions that correspond to the drilling patterns defined by human operators. To achieve this goal, the following specific objectives were defined and accomplished:

(1) To study the state of the art regarding the use of computer vision and deep learning techniques in industrial contexts, with particular focus on segmentation, detection and pattern prediction problems.
(2) To collect and organize a representative dataset of cork strip images, containing both drilled and non-drilled surfaces and to establish the correspondence between paired images captured from different camera perspectives.
(3) To prepare and process the dataset through image processing and deep-learning.
(4) To study and implement a segmentation model, to automatically predict drilling positions that follow human drilling logic.
(5) To analyze the impact of data preparation and pre-training strategies on the performance and generalization ability of the proposed models.
(6) To experimentally evaluate the developed solutions using the constructed dataset, through both quantitative metrics and qualitative analysis of the predicted drilling patterns compared to those of human operators.

1.4. Methodology

To develop this research, the methodology selected was the Design Science Research Methodology (DSRM). This methodology emerged with the aim of filling the lack of a generally accepted process for Design Science (DS) research in Information Systems (IS). Thus, the authors sought to design a model that was consistent with previous literature, that provided a nominal process model to conduct the research and a mental model to present the research output. It offers several benefits in the development of research in IS, providing a structured and rigorous approach for the creation and evaluation of artifacts. The DSRM requires researchers to clearly identify a significant problem and propose innovative solutions, ensuring that the research presents a meaningful contribution to its field. Its emphasis is on the creation of artifacts, which can take the form of constructs (vocabulary and symbols), models (abstractions and representations), methods (algorithms and practices) and instantiations (implemented and prototype systems), which must be rigorously evaluated, demonstrating their quality and effectiveness through well-defined methods, considering not only their functionality but also their impact and contribution to the knowledge base, ensuring their relevance and usefulness. This methodology recognizes that design is an iterative search process, whether new functionalities or requirements are added, or simply to improve performance and quality. There are also different types of approaches to follow, that is, different ways in which researchers can start and conduct the research. These allow the research to begin at different stages, depending on the context and the needs of the project [6]. The different types of approaches are:

- Problem-Centered Approach;
- Objective-Centered Solution;
- Design and Development-Centered Approach;
- Observing a Solution.

In the context of this dissertation, the starting point lies in the identification of a practical problem: the manual process of cork stopper drilling, which is technically demanding, time-consuming and potentially associated with occupational risks, often leading to productivity limitations. This real-world challenge motivated the research and the development of an artifact: a computer vision-based system designed to reproduce the decision-making process traditionally performed by human operators. According to the DSRM process model, this corresponds to the Problem-Centered Approach, where the research begins with a concrete problem that drives the design and evaluation of an innovative solution.

In this work, this approach was applied specifically to address RQ1, which focuses on the development and assessment of an artifact capable of replicating the manual drilling decision process through computer vision techniques.

The stages involved in this methodology follow a nominal sequence, presented in the process model in Figure 1.2.

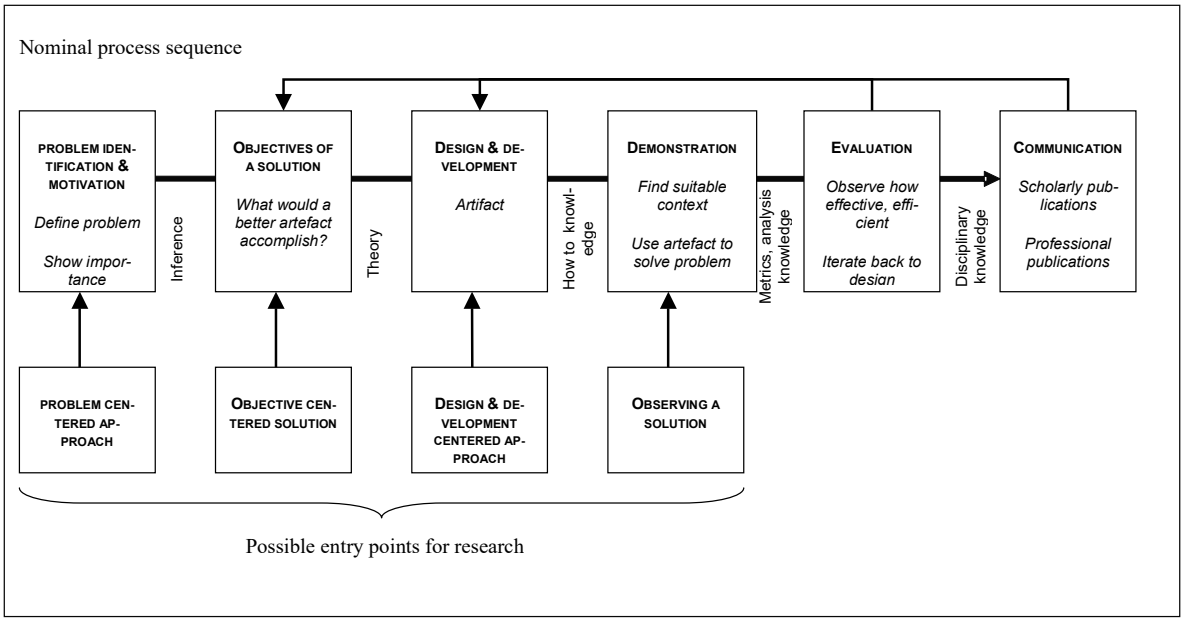


FIGURE 1.2. DSRM Process Model [6]

Thus, the dissertation follows the following sequence:

(1) Problem Identification and Motivation: The research begins with the identification of a practical and relevant problem in the industrial context of cork stopper production. The manual drilling process is technically demanding, time-consuming and potentially associated with occupational risks, often leading to inconsistencies in quality and productivity losses. This motivates the search for an automated, data-driven solution capable of replicating the operator's drilling patterns.
(2) Objectives of the Solution: The main objectives are to investigate whether computer vision techniques can reproduce the drilling decisions made by human operators, to define the features and data representations required for this task and to develop an algorithm capable of predicting optimal drilling positions according to expert patterns.
(3) Design and Development: In this stage, the proposed artifact, a computer vision-based model was designed and implemented. The process involved detecting and segmenting cork strips from captured images, isolating the holes corresponding to real holes, projecting these onto the respective undrilled images, aligning both views and developing the final predictive algorithm using deep learning models.
(4) Demonstration: The functionality of the developed system was demonstrated using real images obtained in an industrial cork stopper production environment, showing the system's ability to predict and visualize drilling positions consistent with human drilling patterns.

(5) Evaluation: The developed artifact was evaluated using the real drilling center coordinates as reference. The comparison between the predicted and real positions allowed assessing the system's accuracy and its ability to reproduce the drilling patterns performed by human operators.
(6) Communication: The results of this work are presented in this dissertation and will be further disseminated through a scientific publication to be submitted after the completion of the research.

1.5. Dissertation Outline

This dissertation is organized into five main chapters, each corresponding to a specific stage of the research process and aligned with the methodology adopted:

- Introduction: provides the contextual foundation of the study. It presents the motivation behind automating the cork stopper drilling process, formulates the research questions and defines the objectives pursued. The chapter also introduces the DSRM adopted as the research framework, outlining the sequence of activities that structure the development of this work.
- Literature Review: analyzes existing research on cork production and the use of computer vision in industrial applications. It summarizes the systematic review process, outlining the databases, selection criteria and filtering stages and highlights the main research gaps that support the relevance of the proposed approach.
- System Design and Development: describes in detail the proposed solution. It begins with a system overview and exploratory analysis of the dataset and proceeds to the design of the vision-based system for the automatic prediction of drilling centers. Each development stage is presented sequentially.
- Results and Discussion: presents the quantitative and qualitative outcomes obtained from the experiments, analyzing the system's performance in predicting drilling patterns. It also discusses the implications, limitations and practical relevance of the proposed approach in real production contexts.
- Conclusions and Future Work: summarizes the main findings of the research, reflects on its scientific and practical contributions and proposes directions for future development.

The schedule shown in Table 1.1 summarizes the temporal organization of the research activities. The process began in December 2024, with the definition of the theme and the initial literature review, followed by the state-of-the-art analysis and the development of the code and models between January and August 2025. The dissertation writing extended throughout the project, ensuring continuous integration between theoretical and practical components while the preparation of the scientific article took place in the final stage of the work.

TABLE 1.1. Dissertation Schedule

Schedule	**2024**	**December**	**January**	**February**	**March**	**April**	**May**	**June**	**July**	**August**	**September**	**October**
Discussion on the theme and scope	X											
Literature review	X		X									
State of the art elaboration		X	X	X								
Code developing				X	X	X	X	X	X	X		
Thesis production			X	X	X	X	X	X	X	X	X	X
Scientific article writing											X	X

CHAPTER 2

Literature Review

In order to ensure a rigorous and structured foundation for the development of this research, Systematic Literature Review (SLR) was conducted, specifically aimed at addressing RQ1. This is a method that follows a systematic and replicable process to collect, analyze and synthesize relevant studies on a topic [7].

Repositories that contain different databases as sources were used, including some with academic interest in computer science.

The following repositories were selected:

- Scopus: https: https://www.scopus.com/
- EBSCO: https://research.ebsco.com/c/2ed5h2/search

To conduct this research, a query was designed. Since it was intended to find literature in Portuguese, English, or Spanish, the keywords were replicated in the three languages. To better illustrate the intent behind the query used in both repositories, it can be divided by the logical operator *AND* into groups as follows:

- The first group (*"computer vision" OR "image processing" OR "machine vision" OR "visual inspection" OR "visão computacional" OR "processamento de imagem" OR "visão de máquina" OR "inspeção visual" OR "visión por computadora" OR "procesamiento de imágenes" OR "visión de máquina" OR "inspección visual"*) defines the technological domain, covering works related to computer vision and automated image analysis in English, Portuguese and Spanish.
- The second group (*"cork" OR "stopper" OR "corkage" OR "rolha" OR "tampa" OR "cortiça" OR "corcho" OR "tapón" OR "taponado"*) specifies the material and industrial focus, cork and its associated products.
- The third group (*"production" OR "manufacturing" OR "fabrication" OR "processing" OR "produção" OR "fabricação" OR "processamento" OR "manufatura" OR "producción" OR "fabricación" OR "procesamiento" OR "manufactura"*) restricts the search to studies related to industrial or manufacturing processes, excluding biological or ecological contexts.
- The fourth group (*"quality control" OR "defect detection" OR "automation" OR "monitoring" OR "controle de qualidade" OR "deteção de defeitos" OR "automação" OR "monitoramento" OR "control de calidad" OR "detección de defectos" OR "automatización" OR "monitoreo"*) focuses on aspects of production automation, defect detection and quality inspection.

Finally, the four groups were combined using the logical operator *AND*, ensuring that only publications simultaneously addressing these four dimensions were retrieved.

The Inclusion Criteria (IC) defined for this query were:

(1) The publication is written in English, Portuguese, or Spanish;
(2) The publication includes the defined keywords in the title or abstract;
(3) The publication must be defined in a relevant area such as Engineering or Information Technologies;
(4) The publication must be open access or belong to the ISCTE-IUL scientific license;
(5) The type of publication is article, paper, or abstract;
(6) The publication addresses an experience related to the production of cork stoppers;
(7) The publication presents the use of computer vision for the production of cork stoppers.

After applying these criteria, duplicates were removed, resulting in a total of 14 related work publications.

Considering that, for this query, no date filters were applied, Table 2.1 makes it possible to understand the temporal dispersion of the publications that relate the use of computer vision with the production of cork stoppers. Most of the publications are more than 10 years old. Considering the evolution of computer vision techniques in recent years, a new contribution to the literature may be essential.

TABLE 2.1. Distribution of Publication Year

Publication Year	Count	Percentage (%)
1994	1	7.14
2001	1	7.14
2002	1	7.14
2005	2	14.29
2006	1	7.14
2007	1	7.14
2008	3	21.43
2010	1	7.14
2015	1	7.14
2021	2	14.29

During the SLR process, it became clear that, in order to replicate the drilling patterns performed by the operator, it would be adequate to understand the types of visual defects in cork that are taken into account by the operator at the moment of drilling. For this purpose, an *ad hoc* search was conducted, not to address RQ1, but rather to meet this specific need. On the other hand, since all studies identified through the SLR were more than four years old, another *ad hoc* search was carried out to find something more state-of-the-art to answer RQ1. In Table 2.2, the distribution of publication years of the studies

identified through the *ad hoc* search can be observed. Compared to the results obtained from the SLR, these publications are considerably more recent, representing more state-of-the-art research that reflects current trends and technological advances in the field.

TABLE 2.2. Distribution of Publication Year *ad hoc* search for RQ1

Publication Year	Count	Percentage (%)
2022	1	20
2023	2	40
2024	2	40

To address RQ2 and RQ3, additional *ad hoc* searches were conducted to complement the findings obtained from the research for RQ1. As a result, one publication from 2025 and another from 2024 were identified, respectively.

Regarding all research questions, a total of 23 related work publications were found.

2.1. Background Concepts

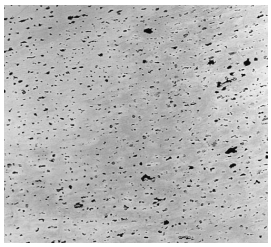
By analyzing the literature found in the research, concepts, definitions and insights related to the problem can be found.

2.1.1. Defects Characterization

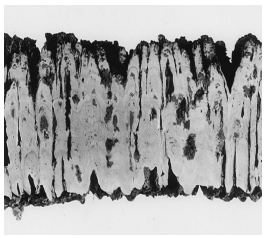
The visual analysis of cork allows detecting some of the defects that it may contain. The evaluation of its quality is carried out at different stages of the cork stopper production process, namely: in the categorization of cork planks (or cork sheets) by quality, before being transformed into strips; and in the final classification of the quality of the finished stoppers. This is an important process, since stoppers of different quality grades assume different economic values. Cork is known to be composed of lenticular channels, which constitute its "normal" porosity, whose coefficients may vary. In [8], the most relevant cork visual defects are identified and characterized as follows:

- "Pore": Lenticular channels that cross the cork planks radially from the inside to the outside, filled with a material of dark or reddish brown colour;
- "Clay": Lenticular channels, usually large, completely filled with powdery material;
- "Bofe": Large lenticular channels that appear as cavities in the cork tissue, without any powder inside;
- "Nail": Inclusions of sclerenchyma in the cork tissue;
- "Cora": Gallery caused by *Coroebus undatus* (F.), always blocked by their excrement;
- "Form": Gallery caused by *Crematogaster scutellaris* (Ol.), always clean and open.

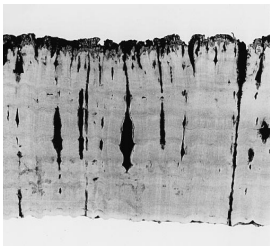
These visual defects are represented in Figure 2.1.



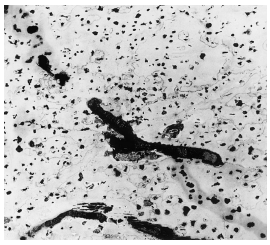
(A) Pore



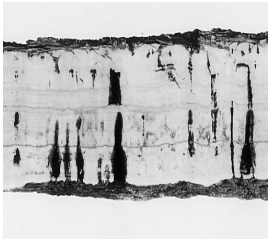
(B) Nail



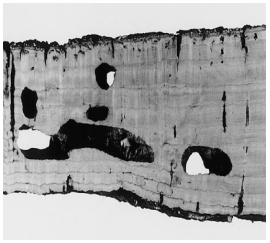
(C) Bofe



(D) Cora



(E) Clay



(F) Form

FIGURE 2.1. Examples of defects [8]

2.1.2. Cork Stoppers Punching

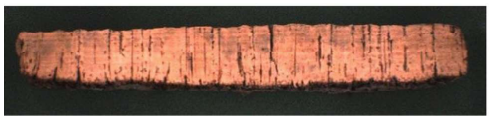
The process of drilling cork stoppers extends beyond the visual detection of defects. To understand this process, it is necessary to understand some of its associated terminology. When the cork sheet is removed from the tree, its inner surface is called the belly (or *barriga*) and the outer surface is called the back (or *costas*), as illustrated in Figure 2.2. While the belly is relatively soft, the back is quite rigid, with cracks and fissures belonging to the visible surface of the tree before the cork sheet is removed. After a few stages of preparation and quality categorization of the cork sheets, they are drilled into smaller units, the cork strips, which are slightly wider than cork stoppers [5].



(A) Back



(B) Belly



(C) Side with belly upwards



(D) Side with belly downwards

FIGURE 2.2. Examples of belly and back [9]

If a cork stopper is drilled close to the back, it is more likely to have defects due to the cracks and fissures typical of this area of the strip. Therefore, drilling should always be performed as close to the belly as possible, where the cork has the best quality. The operator takes all this into account when drilling.

2.1.3. Computer Vision, Object Detection and Image Segmentation

Computer Vision (CV) is a discipline of Artificial Intelligence (AI) in which computers and systems extract valuable insights from digital images, videos and various visual inputs. Thus, it allows machines to observe and understand their environment. CV can take different forms, depending on the type of task to be performed, among which are object detection and image segmentation.

Object detection allows performing different applications, among which locating objects and classifying them by assigning a category. The common models go through three stages: region selection, feature extraction and classification. In the first stage, the relevant regions of the image are identified. Next, the features of these regions are extracted and finally used to classify the objects.

Image segmentation allows extracting the regions corresponding to the objects in the image. Its basis is object detection.

The most widely accepted object detection and image segmentation models are deep learning models. Deep learning consists of the use of multi-layer neural networks, called deep neural networks, whose most representative model is the Convolutional Neural Network (CNN). There is also the combination of regions with the Regions-based Convolutional Neural Network (R-CNN), allowing the model to learn more specific characteristics.

The choice of the model may depend on different aspects, such as accuracy, speed, model size and computational cost [10].

2.2. Related Work

The purpose of this section is to present and discuss the most relevant studies related to the automation of cork processing and the use of computer vision techniques to support or replicate human decision-making during production tasks. By reviewing existing approaches, this section aims to position the present research within the current state of the art, emphasizing methods that involve object detection, image segmentation and predictive modeling applied to cork or other materials with similar irregular textures. The analysis highlights the main methodologies, results and limitations of previous works, identifying the research gaps that motivated the development of the proposed vision-based drilling prediction system.

With regard to the publications related to the cork stopper drilling stage, in [5] we can observe the work most closely aligned with what is intended in this dissertation, that is, obtaining drilling patterns for cork stoppers. In that case, the objective was to develop a prototype of an algorithm capable of generating optimized cork stopper drilling patterns that resulted in less waste compared to the manual method and that also allowed selecting quality grades for the stoppers to be drilled, resulting in different drilling patterns for the same strip. To carry out the experiment, the authors set up an environment to simulate what could be found in a real production context. In the first station, the strip was subjected to laser light in order to detect surface depressions. In the next station, the strip would be manipulated to allow, in the following station, the analysis of the images

obtained by video cameras of the drilling surfaces, enabling the decision of the drilling pattern. Next, the strip would be drilled and ejected. Regarding the image processing algorithm used to reach the ideal drilling pattern, it was divided into three phases:

(1) Generation of the Zone of Acceptable Cork (ZAC): from the grayscale images of the strip drilling surfaces and the belly data obtained by the laser system, the system distinguished between good and bad cork, generating a binary image of the ZAC;
(2) Generation of the Zone of Possible Centers (ZPC): from this image, the system generated a second binary image of the ZPC, that is, the region within the ZAC where it is possible to place the centers of the holes;
(3) Production of an ideal drilling pattern: the vision system, through an algorithm, would then simulate the decision-making process of an operator, in order to select, from the ZPC, the centers that would form an ideal drilling pattern. In addition to emulating the manual process, the system sought to improve it. The algorithm replicated the practice of drilling as close to the belly as possible. Depending on the width of the strip and the drilling diameter, it might be possible to perform a second pass to use the remaining material. The first center (which influences all the others) would be chosen by calculating the perpendicular distance of a set of possible centers to an imaginary line that passes through the upper corner of the, at $30^{\underline{o}}$ relative to the vertical y-axis. The first center would be the closest to the line, located in the upper-left region of the ZPC. Iteratively, the ZPC would be reduced through erosion at the chosen center, with a structuring element twice the drilling diameter. When the ZPC was eliminated, the drilling process would stop. The system also had three operating modes, which would result in cork stoppers of different qualities, related to the presence of defects in the drilling regions. The tests carried out demonstrated a production capacity equivalent to that of two operators, being considered positive, although there was no quantification of these results, which would make them more reliable. Since the publication dates from 1994 and the algorithm is based on simple image processing, its scalability for application to strips of very different shapes from those tested may be questionable.

An implementation of an autonomous system for feeding cork strips into a drilling machine was proposed in [9]. The approach was based on the traditional Portuguese classification of cork surfaces into *barriga* (belly) and *costas* (back), which correspond to distinct textural and mechanical properties relevant for determining the drilling side. The system was assembled using a conveyor belt, onto which cork strips were placed and analyzed using computer vision techniques. To isolate each strip from the background, the authors employed background subtraction provided by the OpenCV library [11], enabling effective segmentation of the strip's contour. Since the location and orientation of the holes are directly related to the strip's quality (holes must be drilled close to the *barriga*

to optimize the cork yield and reduce waste) the system required precise spatial alignment before feeding the strip into the machine. For this purpose, a CNN was trained to classify the orientation and position of each strip. When necessary, the CNN would instruct a robotic arm to rotate the strip before insertion, ensuring correct alignment relative to the drilling mechanism. The authors constructed training, validation and test datasets for this task and achieved an impressive classification accuracy of 100%, demonstrating the feasibility of using computer vision and deep learning to automate tasks traditionally performed manually in cork processing.

Combining computer vision and robotic manipulation, a fully automated system for feeding cork strips into drilling machines was presented in [12]. The system was designed to address one of the main bottlenecks in the cork stopper production line: the manual feeding of strips into automatic drillers. The proposed setup included a conveyor belt, a SCARA robotic arm, a custom pneumatic gripper and a vision module responsible for detecting, segmenting and classifying the cork strips. Using a camera placed above the belt, the system performed background subtraction (based on the Mixture of Gaussians model available in OpenCV) to isolate the cork strips and extract geometric information such as position and orientation. A machine learning classifier, specifically a Support Vector Machine (SVM) trained with texture features derived from the Gray-Level Co-Occurrence Matrix (GLCM), was employed to identify which surface of the strip was visible (*barriga*, *costas* or *side*). The correct identification of the visible side was crucial for orienting the strip before feeding it to the driller, ensuring that the best quality surface faced the machine. Experimental results demonstrated 100% classification accuracy on a test dataset of over one hundred cork strips, showing the robustness of the proposed method under controlled lighting conditions. Although the manipulation stage was only partially validated due to logistical constraints, the study provided an important contribution to the automation of cork feeding processes and demonstrated the potential of combining traditional vision algorithms with machine learning for real-time industrial applications.

In [13], the authors extended their previous work by addressing the problem of segmentation and robotic manipulation of cork strips stored in bulk. Unlike earlier approaches, which assumed that the strips were already aligned on a conveyor, this study focused on the initial stage of handling randomly piled strips, aiming to increase the autonomy and efficiency of the feeding process. The authors proposed an integrated system combining 3D computer vision and robotic grasping, designed to detect, localize and extract individual strips from an unorganized pile. The segmentation stage relied on data acquired from a stereo vision system, which produced dense point clouds representing the 3D structure of the cork pile. A Conditional Euclidean Clustering algorithm was then applied to these data, grouping points into clusters based not only on their Euclidean distance but also on the similarity of their surface normals. This conditional criterion allowed the system to effectively separate cork strips even when they were partially in contact, as

long as there was a sufficient difference in depth or surface orientation. The resulting clusters were used to estimate the position and orientation of each strip, enabling subsequent manipulation. A 6-DOF robotic manipulator equipped with a custom pneumatic gripper was then employed to pick and place the strips onto a conveyor belt for downstream processing. Experimental results demonstrated a success rate of approximately 90% in correctly identifying and grasping isolated strips under controlled conditions. However, performance degraded when strips were heavily overlapped or when low-texture regions reduced stereo reconstruction quality.

Publications [14–24] are related to other stages of the cork stopper production process, focused entirely on quality classification, whether of cork planks or cork stoppers.

One of the earliest applications of computer vision to cork quality assessment was proposed in [14], a system to classify cork planks based on surface texture and color features. The approach used image analysis techniques combined with statistical feature extraction, including gray-level histograms and co-occurrence matrices, to characterize surface defects and variations. A supervised classification model was then trained to distinguish between quality grades of cork. The study demonstrated the feasibility of automating visual inspection, achieving high accuracy under controlled lighting conditions, although its applicability was limited to planar surfaces and did not address process automation.

An automatic vision system for real-time cork stopper classification was presented in [15], aiming to standardize the grading process and reduce subjectivity in manual inspection. The proposed modular system combined a mechanical setup with a linear CCD camera and a networked acquisition–processing architecture, enabling high-speed image capture and analysis. Image processing involved morphological operations, convolution, binarization and clustering to detect and quantify surface defects such as cracks and holes, while the classification stage assigned each stopper to one of eight quality classes according to defect size and area. The system achieved stable and repeatable results under controlled conditions.

Aiming to reduce the subjectivity and cost of manual inspection, the authors in [16] proposed a computer vision system for cork stopper quality classification. The study evaluated several feature extraction techniques, including global and local thresholding methods, texture analysis based on gray-level co-occurrence matrices and defect size estimation through morphological operations. Euclidean distance, k-means, k-nearest neighbors and a neural network were compared, with the Euclidean classifier using selected features (defect area, texture contrast, texture entropy and largest defect size) achieving the best classification results. The system reached an average classification error of 45%.

A real-time computer vision system for cork stopper inspection and classification based on Bayesian decision theory was described in [17]. Each stopper was represented by 43 extracted visual features, including defect area, number of pores, gray-level statistics and surface texture descriptors. Several parametric and non-parametric classifiers were tested, with the Class-Conditional Independent Component Analysis (CC-ICA) approach

achieving the best performance. This method improved probability density estimation in high-dimensional feature spaces, reaching 98% classification accuracy across five quality grades.

A comparative study of classification algorithms for cork stopper and disk quality evaluation was proposed in [18]. The authors tested five approaches: euclidean distance, k-means, k-nearest neighbors, a classical neural network and a Neuro-Fuzzy classifier. They used a dataset of 700 images across seven quality classes. Feature extraction included defect area, largest defect size and texture descriptors derived from Laws' texture energy measures. The Neuro-Fuzzy model achieved the best performance, with an error rate of 4.28%, significantly outperforming traditional classifiers.

A real-time system for cork stopper quality classification implemented on FPGA hardware was presented in [19]. The proposed architecture combined a Visual C++ front-end with a reconfigurable HOT2-XL PCI board, allowing several image processing operations including thresholding, filtering and morphological transformations to be executed directly in hardware. The system extracted geometric features such as defect area, largest defect size and defect-to-surface ratios, classifying stoppers according to user-defined quality thresholds.

A comparative study to determine the most suitable classifier for automatic cork quality evaluation was proposed in [20]. Using a dataset of 700 cork disk images across seven quality grades, the authors extracted features such as defect area, texture contrast, entropy and largest defect size, obtained through thresholding and texture analysis. Four classifiers, such as Backpropagation neural network, k-means, k-nearest neighbors and Euclidean distance were tested, with the Euclidean classifier achieving the best performance and a classification error of 61.4%

A neuro-fuzzy system for automatic cork quality classification, combining image processing with fuzzy logic and neural networks to handle class overlapping was described in [21]. Using the Cork Image Database (CID) with 700 cork disk images divided into seven quality grades, six discriminative features were extracted, including defect area, largest defect size and Laws' texture energy measures. The fuzzy membership function enabled modeling of uncertain classifications between adjacent grades. The proposed $6\times33\times7$ network architecture achieved an error rate of 28.28%, significantly improving over previous studies.

For industrial cork stopper quality classification, the authors in [22] presented an advanced Neuro-Fuzzy vision system. Using a dataset of 700 cork disk images across seven quality classes, the authors evaluated several feature extraction techniques, including global and local thresholding, co-occurrence texture matrices, Laws' Texture Energy Measures and heuristic features such as hole detection and largest defect size. The selected six best features were combined and tested with five classifiers. The Neuro-Fuzzy classifier achieved the best performance, reaching an error rate of 4.28%.

A computer-vision system for automatic cork stopper classification based on quantitative porosity analysis was presented in [23]. Using image processing to detect pores on the body and tops of stoppers, the authors applied Canonical Discriminant Analysis (CDA) and Stepwise Discriminant Analysis (SDA) to identify the most discriminative features, such as primarily porosity coefficient, maximum pore dimensions and number of pores. Decision rules with defined threshold limits were then established to assign stoppers to one of seven quality classes. The method achieved an overall class-yield match of 86%.

Image analysis techniques were applied in [24] to characterize the porosity of wine cork stoppers across seven commercial quality classes. Using high-resolution images of both the lateral and top surfaces, features such as pore area, width, length, porosity coefficient and number of pores per 100 cm^2 were extracted and statistically analyzed. The study found a clear trend of increasing porosity and pore size from higher to lower quality grades, although significant overlap among mid-quality classes limited full discrimination. Reducing the categories to three quality levels (superior, medium and inferior) improved statistical separability, demonstrating that porosity coefficient and pore count could serve as reliable indicators for future standardized grading systems.

Publications [25–29] are the result of an *ad hoc* search conducted to complement the studies identified through the SLR, focusing on more recent and state-of-the-art approaches.

An up-to-date contribution to the application of computer vision and deep learning techniques in the cork industry was described in [25], expanding the field beyond the traditional focus on the quality inspection of cork stoppers. They address the problem of automatic quality detection of cork discs, which are circular pieces of cork obtained from oak bark and used in the manufacturing of badminton shuttlecock heads. Despite being a different industrial product, the visual and structural characteristics of these cork discs are closely related to those of cork stoppers and cork strips used in other production contexts, such as the one explored in the present dissertation. The article emphasizes that cork is a highly heterogeneous natural material, whose texture, color and surface defects are strongly affected by environmental conditions during growth. As a result, it is nearly impossible to find two cork samples with identical patterns of pores or imperfections. This intrinsic variability makes manual visual inspection unreliable, as it depends heavily on the operator's experience, visual fatigue and subjective judgment. To overcome these limitations, the authors propose an automated optical inspection system combining industrial imaging and deep-learning-based object detection to ensure more consistent and efficient quality assessment. They developed a dataset of over 8500 high-resolution images of cork discs, photographed on both sides with industrial cameras and annotated for defect types (holes, notches and black spots) and quality grades (qualified, unqualified, outer bark). The system combines defect detection and quality classification, linking visual features to manufacturing decisions. The approach builds upon the YOLOv5 architecture, enhanced with several innovations: a Convolutional Block Attention Module to highlight subtle

texture variations, a Center-Match and Shortest-Distance Label Assignment strategy for improved sample selection and a Detection Result Processing algorithm to refine classification based on detected defects. To compensate for the lack of defective samples, the authors also employ a Generative Adversarial Network (GAN) based data augmentation method that synthesizes realistic defects. Experiments conducted achieved a mean F1-Score of 86.7%. The results achieved demonstrate that deep learning can provide fast, precise and reliable inspection of cork surfaces, suitable for industrial deployment. This study is particularly relevant to the current research as it confirms that deep-learning vision models can effectively learn complex visual patterns in cork, despite its natural heterogeneity.

One of the first deep-learning based approaches for the automatic visual classification of cork stoppers was presented in [26], replacing the traditional manual inspection process still predominant in cork manufacturing. The authors proposed a Multi-Scale Faster R-CNN with Upper Lower Layer Fusion (UMS-FRCNN) to address the difficulty of detecting fine-grained texture details and small defects on cork surfaces, which vary significantly in color, porosity and shape due to their natural origin. The network combines high-level and low-level convolutional feature maps to enhance the detection of small regions of interest and integrates a GAN classifier to improve multi-class discrimination among six industrial quality grades (4A–C). The system captures multiple side-view images of each cork using a camera-mirror setup and classifies them based on surface defects such as mildew, black spots, red dots, holes, bark residues or incomplete areas. Experiments conducted on 1 000 cork samples demonstrated the model's superiority over traditional detectors such as RetinaNet, YOLO v3 and standard Faster R-CNN. After 2 000 training epochs, the proposed method achieved 89.8% accuracy, 87.4% recall and an mAP of 0.93, while also processing 25 corks per second, making it suitable for real-time factory deployment. An ablation study confirmed that the upper and lower layer feature fusion substantially improved performance, particularly for small defect detection. The proposed system has been successfully implemented in an industrial cork-sorting plant, proving the feasibility of deep-learning-based quality inspection in cork production and marking a significant shift from manual, subjective evaluation to automated, data-driven classification.

A deep-learning based approach to automate the estimation of cork volume in cork oak trees was introduced in [27], addressing a key problem in forest management. Traditionally, cork yield estimation relies on field measurements of tree height and diameter, a process that is labor-intensive, time-consuming and limited in precision. The authors proposed a three-stage pipeline combining Mask R-CNN for instance segmentation with machine learning regression models for cork volume estimation. A custom dataset of 62 field images was created, where reference targets of known size were affixed to tree trunks to calibrate area measurements. The Mask R-CNN model was trained to segment both the trunk and the reference targets, achieving a mean Average Precision (mAP@0.7) of

0.96, demonstrating high segmentation accuracy even with limited data. From the predicted trunk masks, the corresponding areas were calculated and converted into metric units using the target references. These measurements, combined with biometric features of the trees, were then used to train three regression models: Linear Regression, Support Vector Regression (SVR) and a Multilayer Perceptron. The SVR model achieved the best results, with an average volume estimation error of 8.75%, comparable to traditional field-based methods.

A strategy to enhance the robustness of computer vision models for manufacturing defect detection was proposed in [28], addressing the problem of poor generalization due to repetitive, homogeneous training data. The authors compare classification and object detection approaches using ResNet-50 and RetinaNet backbones, trained on datasets of metal parts containing artificial defects. Their experiments show that models trained on diverse backgrounds generalize better to unseen data, while object detection models outperform simple classifiers in terms of AUC, especially when tasked with identifying defect locations. The study also employs clustering to identify which types of training images contribute most to generalization, concluding that combining diverse data with object detection leads to more reliable inspection systems adaptable to new production conditions.

In [29], the authors proposed a real-time defect detection framework for pharmaceutical blister packaging based on an improved version of YOLOv8, named CBS-YOLOv8. The study addresses the limitations of manual visual inspection and traditional image-processing methods, which are slow, error-prone and unsuitable for continuous production lines. The proposed model integrates three key enhancements: Coordinate Attention (CA) to strengthen spatial feature extraction and preserve positional information, Bi-Directional Feature Pyramid Network (BiFPN) to improve multi-scale feature fusion and SimSPPF to reduce computational overhead and increase inference speed. The system was trained on a custom dataset of defective tablets (including broken, cracked, empty, color-mismatched and contaminated samples) and evaluated under real production conditions. Experimental results demonstrated that CBS-YOLOv8 achieved a mean Average Precision (mAP@50) of 97.4%, an F1-score of 97% and a real-time inference speed of 79.25 FPS, outperforming YOLOv5, YOLOv7 and standard YOLOv8 variants. The model was also validated on the SESOVERA-ST saline bottle dataset, achieving 99.3% mAP@50, confirming its generalization capacity across industrial inspection tasks. The proposed method combines high detection accuracy with low latency, making it suitable for real-time quality assurance in manufacturing. Its design principles, feature coordination, efficient fusion and computational optimization, reflect the current evolution of deep-learning based visual inspection systems toward Industry 4.0 and autonomous production monitoring.

In addition to the systematic and *ad hoc* research conducted for RQ1, an *ad hoc* search was performed for RQ2, namely to understand how data preparation and pre-training strategies influence model performance and generalization in computer vision tasks.

The impact of pre-training strategies on the performance of deep-learning models for industrial defect detection was investigated in [30], focusing on how domain-specific data can improve generalization and convergence. The study compared three training configurations: random initialization, ADE20K based pre-training and pre-training using a custom industrial defect dataset, across eight manufacturing domains comprising over 12,000 annotated images with various defect types such as cracks, scratches, corrosion and oxidation. Using a U-Net architecture, the author demonstrated that models pre-trained on industrial domain data achieved consistently higher Intersection Over Union (IOU) scores and faster convergence compared to both randomly initialized and ADE pre-trained models. Specifically, the industrially pre-trained networks improved IOU by up to 3–5% across different defect datasets, while maintaining lower escape rates (missed defects) despite a slight increase in overkill rates (false positives). The author also emphasized several key data preparation principles for successful transfer learning in manufacturing contexts, including class balance, exclusion of ambiguous labels and use of cosine annealing for learning-rate scheduling. These findings highlight that domain aligned pre-training and well curated defect datasets are critical for achieving robust and generalizable vision systems in industrial environments.

In order to complement the discussion of RQ3, an *ad hoc* search was conducted to investigate whether segmentation models can be effectively adapted to predict drilling patterns on irregular or textured materials while maintaining robustness and consistency under diverse visual and structural conditions.

In [31], the authors developed a deep-learning based approach for the automated segmentation of drilled holes in textured wooden furniture panels, aiming to overcome the limitations of traditional image-processing methods in visually complex industrial environments. The authors proposed several lightweight U-Net encoder–decoder architectures, integrating residual connections, Atrous Spatial Pyramid Pooling (ASPP), squeeze-and-excitation modules and CoordConv layers to enhance contextual feature extraction and robustness to lighting and texture variations. A dataset of industrial-grade MDF panels exhibiting diverse textures, colors and surface defects was acquired using a line-scan camera system and extensive data augmentation was applied to simulate real production variability. Experimental results demonstrated that the enhanced U-Net models significantly outperformed conventional methods (Sobel, Laplace, Canny) and the baseline U-Net, achieving a Dice score of 97.89%. The best configuration combined residual connections, ASPP and squeeze and excitation blocks, providing more stable predictions across heterogeneous textures without substantial computational overhead. These findings show that context aware encoder–decoder architectures can achieve accurate and

consistent segmentation of drilling patterns even under complex visual and structural conditions.

2.3. Contribution

Overall, the reviewed literature reveals a progressive evolution of computer vision applications in cork processing, from early rule-based and statistical image analysis methods to the recent integration of deep-learning architectures capable of handling the material's natural variability. The earliest works predominantly relied on handcrafted features, simple texture descriptors and fuzzy or statistical classifiers, achieving acceptable results under controlled conditions but with limited scalability to diverse visual contexts or real-time operation. These studies established a solid foundation by formalizing the challenges inherent to cork, namely its irregular texture, uneven illumination and high intra-class variability. In contrast, more recent research demonstrates a clear shift toward data-driven and learning based approaches, where convolutional and transformer based models have started to dominate industrial inspection tasks. Within the cork industry, these advances have materialized mainly in quality inspection and surface defect detection, as illustrated by the studies [25] and [26]. These works introduced architectures such as YOLOv5 and multi-scale Faster R-CNN, highlighting the importance of attention mechanisms, multi-level feature fusion and data augmentation to improve defect localization in natural materials. Other domains, such as metal and composite manufacturing, have contributed complementary insights, particularly regarding pre-training and data preparation strategies [30] that enhance model generalization across variable lighting and texture conditions. Furthermore, research like that of [31] demonstrates that segmentation models, such as U-Net, can accurately delineate drilled holes in complex textured materials, confirming their suitability for tasks that require spatial precision and robustness. Together, these findings suggest a convergence toward context-aware, multi-scale deep networks as the state-of-the-art paradigm for visual inspection and predictive modeling in heterogeneous materials.

In summary, the literature reviewed provides valuable answers to the formulated research questions:

- Regarding RQ1, it confirms that automation in cork processing has traditionally focused on segmentation and classification stages, with limited work addressing predictive decision-making.
- Concerning RQ2, recent studies show that pre-training on domain-specific data and careful data preparation play a crucial role in improving model convergence and robustness.
- Finally, for RQ3, current trends highlight the effectiveness of adapting segmentation models, such as U-net, for achieving structural consistency and robustness under varying visual and textural conditions.

Nonetheless, there remains a research gap in systems that move beyond inspection and classification to predict drilling patterns, that is, learning to replicate the operator's spatial decision-making process. The present dissertation aims to bridge this gap by developing a vision-based drilling prediction system that leverages these recent advances in deep-learning architecture design, pre-training and data preparation.

CHAPTER 3

System Design and Development

3.1. System Overview

The system developed in this work aims to automate the identification of drilling locations on cork strips for the production of natural cork stoppers, a task that is currently carried out manually by experienced human operators. In the traditional process, the operator visually inspects each cork strip, identifying the regions where holes can be drilled without intersecting defects. This decision-making process relies heavily on human expertise and visual perception, which makes it difficult to replicate consistently and objectively across production lines. The proposed system seeks to reproduce this process through computer vision and machine learning techniques. Given a set of high-resolution images of cork strips, the system is designed to output the coordinates of the drilling centers. These coordinates represent the predicted positions where the holes should be made. In this sense, the system functions as a vision-based decision-support module that can be integrated into a fully automated drilling environment. From an industrial perspective, the predicted drilling coordinates can be directly transmitted to an automatic drilling system, such as a robotic arm equipped with precision actuators. By doing so, the entire drilling stage could operate autonomously, reducing human dependency while ensuring consistency and repeatability in production.

The high-resolution images used as the systems' input were acquired in a real cork stopper production environment. In the industrial setup where the dataset was collected, each cork strip is placed on a fixed table and simultaneously captured by two cameras positioned on opposite sides: the left and right cameras, as shown in Figure 3.1. Both cameras photograph the strip's surfaces (left and right), depending on its orientation. In order to determine the visual patterns that resulted in drilling at specific locations, each strip was photographed both before and after the drilling operation. In the post-drilling images, the holes appear as dark circular areas, while the undrilled images show the natural texture and possible defects of the cork surface.

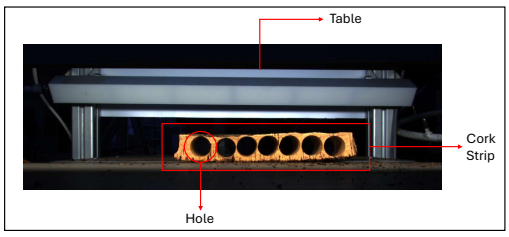


FIGURE 3.1. Left view of the data acquisition setup. The cork strip is positioned on the table between the two cameras.

Figure 3.2 represents the top view of the experimental setup. The two cameras are aligned symmetrically along the strip's longitudinal axis, ensuring a consistent perspective between both sides. This configuration is important because defects can appear only on one surface (or in both), so both must be analyzed. The cameras are synchronized, meaning each pair of images correspond to the same moment in time, which is essential for accurate alignment between the two sides. The grey lines represent the field of view of each camera, illustrating the overlapping region captured by both perspectives.

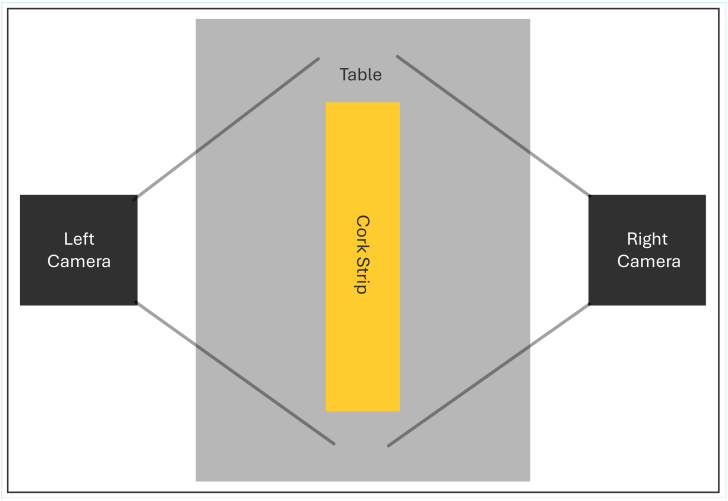


FIGURE 3.2. Representation of the top view of the acquisition setup showing the relative position of the left and right cameras, the table and the strip.

To facilitate the understanding of the designed system's overall behavior and workflow, a Business Process Model and Notation (BPMN) diagram was developed, as shown in Figure 3.3. BPMN is a standardized graphical language used to represent processes in a clear and intuitive way, illustrating both it's starting and ending points, as well as the key intermediate stages that transform the input images into the predicted drilling coordinates. By presenting the process in this way, it becomes easier to grasp the logical flow and dependencies between each component of the system before examining their technical details in the following sections.

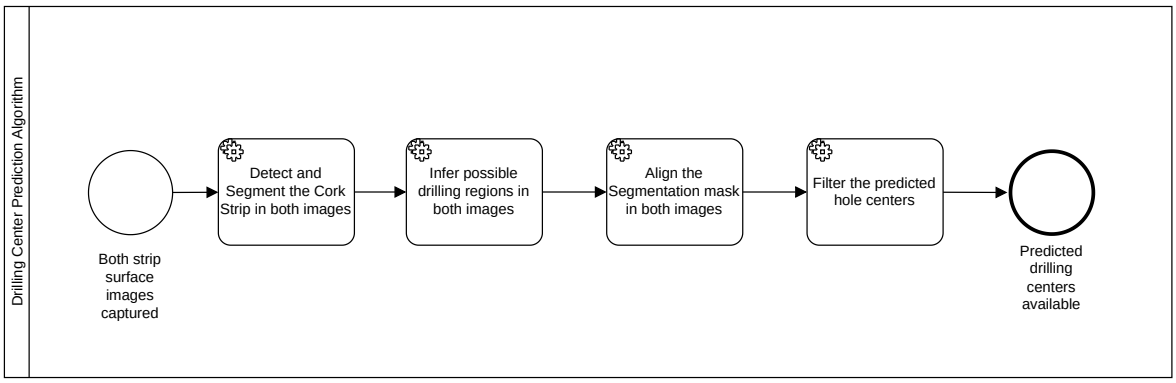


FIGURE 3.3. BPMN diagram of the System's workflow

The process begins when the system is provided with the images corresponding to a single cork strip in its undrilled state. Each subsequent task is sequential and dependent on the previous ones, as illustrated in the diagram. The system first detects and segments the strip in each image. Detection refers to locating the strip within the image, while segmentation refers to extracting its exact contour, resulting in a segmentation mask that isolates the strip from the background, an essential step for all subsequent operations.

Next, the system analyzes each surface of the strip to identify the regions that appear suitable for drilling, simulating the visual inspection performed by the human operator. Since both surfaces may contain defects (or defects may exist only on one of them) it is necessary to align the two surfaces so that every region on one side corresponds precisely to the same region on the opposite side.

This requirement can be better understood by considering that, when a hole is drilled, it passes through the entire width of the strip. Therefore, both surfaces are affected by each drilling operationa nd it is crucial to ensure that a hole made on one side does not intersect a defect on the other side. This visual reasoning is also part of the operator's manual decision-making process.

Finally, after identifying the regions suitable for drilling on both surfaces and considering that one surface (for instance, the left one) is drilled first (from the side where the drill bit is positioned) the system must ensure that each predicted hole will not reach defects on the opposite surface. Consequently, only those holes predicted on the left surface whose corresponding locations are also valid on the right surface are accepted.

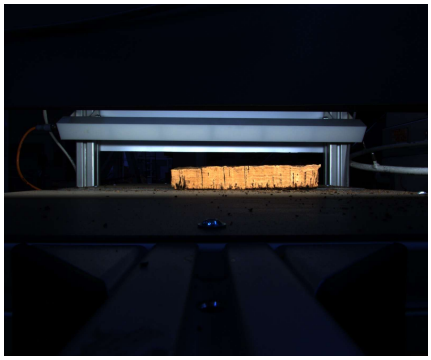
The process concludes with the system outputting the coordinates of the predicted drilling centers for the strip.

3.2. Dataset Analysis

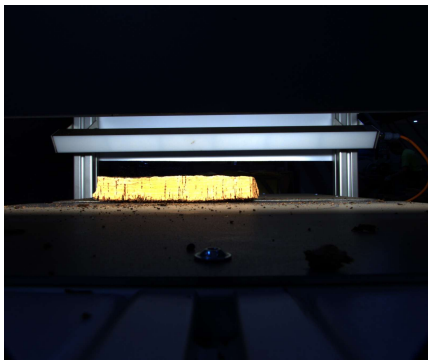
This section describes the structure and organization of the dataset used in the development of the proposed system. While the Section 3.1 detailed how the images were acquired and the experimental setup in which the data collection took place, the focus here is on how the dataset was organized to facilitate processing and model training.

Each cork strip in the dataset is represented by four images: a left view before drilling, a left view after drilling, a right view before drilling and a right view after drilling. This structure allows the system to analyze both surfaces of the same strip and to correlate the visual appearance before and after the drilling process.

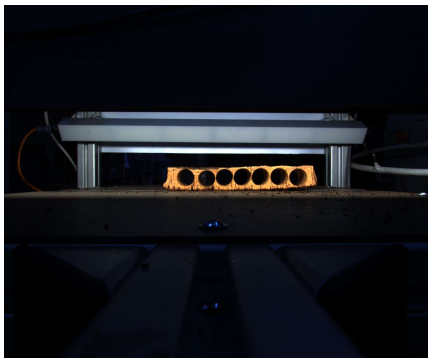
The dataset is organized chronologically according to the timestamps embedded in the filenames. Figure 3.4 shows examples of images that correspond to the same cork strip: 20240313090031_L, 20240313090046_L, 20240313090031_R and 20240313090046_R. When the files are sorted in ascending order by name, the four images belonging to each strip naturally group together, enabling a consistent pairing of left and right views, as well as drilled and undrilled states.



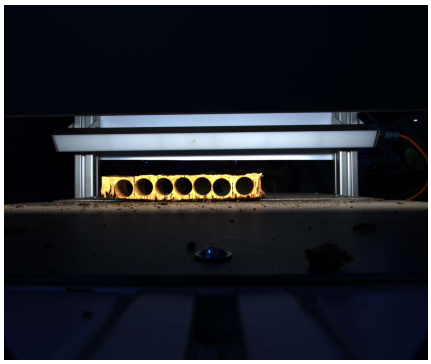
(A) 20240313090031_L: Left undrilled view



(B) 20240313090031_R: Right undrilled view



(C) 20240313090046_L: Left drilled view



(D) 20240313090046_R: Right drilled view

FIGURE 3.4. Example of the four views of a cork strip

The dataset used in this work was obtained from a larger image collection made available by an industrial cork processing company. Table 3.1 shows the selected dataset characterization. From the larger image collection provided by the cork processing company, a total of 5348 images were selected for this work. Among these, 116 images contained only background (without any cork strip). The remaining 5232 images correspond to samples that effectively contain cork strips. Considering that each strip is represented by four images (left and right views, before and after drilling), the final dataset comprises 1308 unique cork strips.

TABLE 3.1. Summary of the selected dataset

Description	**Count**
Total images selected from source dataset	5,348
Images containing only background	116
Valid images containing cork strips	5,232
Unique cork strips represented	1,308

3.3. System Development Stages

The main objective of this work is to develop a computer vision system capable of predicting the drilling center locations on cork strips, that is, automatically determining where

holes should be made, in a way that replicates the decisions normally taken by human operators.

In this context, the predictive model refers to a supervised learning algorithm that must be trained using images where the correct hole positions (the ground truth) are known. These positions are typically provided as annotations: manually or automatically defined markers indicating the exact locations of the holes in each image. However, the original dataset provided by the cork processing company did not include such annotations, meaning that the hole locations were not recorded for the undrilled strips. As a result, the direct training of a predictive model on these images was not feasible. Furthermore, the visual differences between the images of the same strip before and after drilling, such as small strip displacements, illumination variations or surface deformations caused by the drilling process, introduce additional challenges that must be addressed prior to model training.

To overcome these limitations, several system development stages were implemented to process and structure the available data, enabling the definition of training labels and the preparation of the images for prediction. These stages are represented in Figure 3.5 and include detecting and segmenting the cork strips, identifying and segmenting the holes on the drilled images, projecting the hole coordinates onto the corresponding undrilled strips and generating the ground-truth data required for model training.

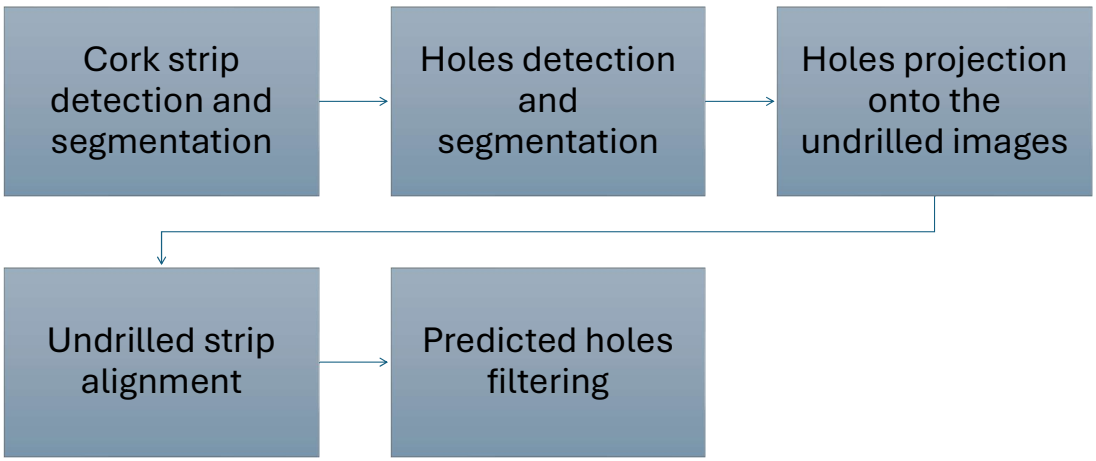


FIGURE 3.5. Overview of the system development stages

In addition, other stages such as the alignment of the undrilled images of the same strip and the filtering of the predicted holes are also addressed. Unlike the previously mentioned stages, these do not prepare the data for model training. Instead, they are applied after the model predictions have been generated, in order to further refine the results and ensure that the predicted drilling locations better replicate the logical decision process of the human operator. Each of these stages is individually described and discussed in the following subsections.

3.3.1. Cork Strip Detection and Segmentation

Considering the original images (shown in Figure 3.4), the first step to prepare the images for the model training consisted of detecting the cork strips, that is, locating each strip within the image in order to enable the subsequent processing stages,

At an initial stage, a simpler image-processing approach was tested using the OpenCV library [9]. Based on the work presented in [11], the background subtraction method was explored to segment the cork strips. This method consists of subtracting a background image, pixel by pixel, from another image, highlighting the regions with significant differences. It is a simple technique commonly used to detect movement or segment objects when a fixed background is available. OpenCV provides several variants of this method, each suited to different scenarios:

- Geometric Mixture of Gaussians (GMG), capable of adapting to lighting variations;
- Mixture of Gaussians (MOG), adaptable to slight background changes;
- MOG2, an evolution of MOG, better suited to complex environments (e.g., with shadows).

Both strip images (left and right views) and corresponding background images (also for both views) were used. All three variants were tested and compared. It was found that the MOG variant produced the most suitable results when applied directly, without additional preprocessing. This approach generally succeeded in distinguishing the cork strip region, demonstrating that the model could "understand" the common background pattern. However, frequent errors were observed in the resulting binary segmentation masks: in some images, small cork debris and, occasionally, parts of the table were incorrectly included.

Figure 3.6 shows the results of applying the background subtraction algorithm to four different image examples. The background regions are represented in black, while the white areas correspond to the regions identified by the model as the difference between the background image and the image containing the cork strip. In sample D, for instance, it can be observed that the model incorrectly included part of the table in the result, which could negatively affect the performance of the subsequent stages of the system.



(A) Example 1 (B) Example 2

(C) Example 3 (D) Example 4

FIGURE 3.6. Examples of binary masks resulting from the application of background subtraction (MOG)

To mitigate these issues, additional OpenCV morphological operations were applied, such as identifying the largest connected component (to remove small noise such as debris) and dilation (to fill "holes" in the masks). The resulting masks appeared cleaner, with much less debris. Nevertheless, due to the irregular nature of cork, characterized by cavities and texture variations, the method tended to misclassify some of these regions as background, including portions of real holes. On the contrary, parts of the table sometimes remained included in the masks and their removal compromised full detection of the strip.

Figure 3.7 shows the intersection of the masks obtained from the improved background subtraction method with the corresponding real images from Figure 3.6, allowing the true texture of the cork strips to be visualized. It can be observed that, for example, the segmentation in sample 4 has improved compared to the previously shown result, with part of the noise successfully removed from the mask. However, a portion of the table surface is still visible in the same image.



(A) Example 1 (B) Example 2

(C) Example 3 (D) Example 4

FIGURE 3.7. Examples of intersection between the masks resulting from the improved application of the background subtraction method and the real images

To improve upon the limitations of classical image processing, a more advanced deep-learning-based object-detection method was tested: the YOLO version 11 model, available in the Ultralytics package [32]. YOLO is one of the most efficient and widely used architectures for both object detection and real-time image segmentation. In its default form, it is pre-trained on the COCO dataset, containing 80 generic object classes. Since cork strips are not among these classes, supervised training was required. For training, a dedicated dataset was prepared containing:

- a folder with the selected images;
- a folder with their corresponding annotations;
- a YAML configuration file describing the paths to images and labels, as well as the number and names of the classes.

Following YOLO's standard object-detection format, each instance was annotated on a single line as `<class_id> <x_center> <y_center> <width> <height>`, where `class_id` is the integer identifier (starting at 0) and the coordinates are normalized to the image *width* and *height* (values in [0, 1]). `<x_center>` and `<y_center>` correspond to the center of the bounding box.

Annotations were generated using Labelme [33], by manually drawing bounding boxes in 341 images. The coordinates of these boxes were then converted into YOLO-compatible annotations.

Model performance was evaluated using the mAP 50–95 metric, which is the mean of the average precisions computed across IOU thresholds from 0.5 to 0.95 [34]. A value of 0.94 mAP 50–95 was achieved.

This model successfully extracted bounding boxes for all cork strips in the dataset, defining the regions of interest for the following processing stages.

Figure 3.8 shows examples of cork strip detections obtained using the YOLO model. The blue rectangle represents the *bounding box*, a rectangular region that encloses the detected object and defines its spatial extent within the image. The value displayed above each box corresponds to the model's confidence in the prediction, that is, the probability that the detected region actually contains the object. In examples 1 and 2, high confidence values of 0.95 and 0.94 were achieved, respectively, indicating that the model was highly certain about the detections.



(A) Example 1 (B) Example 2

FIGURE 3.8. Examples of cork strip detections by the YOLO model

Since YOLO can also perform instance segmentation, the extracted bounding boxes were reused to generate binary masks for training its segmentation variant. To achieve this, an adaptive-threshold binarization followed by morphological operations and contour extraction was applied to create polygons representing the strip area with minimal noise. This approach favored clean masks, even if they slightly under-segmented the object. Annotations were converted into YOLO's segmentation format, in which each line consists of the `<class_id>` followed by a sequence of normalized vertex coordinates `<x_i> <y_i>` defining the object polygon.

The trained segmentation model reached values of mAP 50–95 equal to 0.762. However, a visual inspection revealed that the segmentation masks did not accurately match the true shape of the cork strips. Figure 3.9 shows examples of image segmentations obtained using the YOLO model. In each example, the image is displayed in black except for the region corresponding to the *bounding box* previously extracted by the YOLO segmentation model. It can be observed that the model fails to properly fit the contours of the object, resulting in masks that only roughly approximate the actual shape of the strip.



(A) Example 1 (B) Example 2

FIGURE 3.9. Examples of image segmentation with the YOLO model

To achieve greater robustness, another model was tested, SAM, version 2 [35], developed by Meta AI, also available in the Ultralytics package. SAM can segment virtually any type of image without prior class-specific training, being therefore considered a zero-shot segmentation model. According to its documentation, SAM can segment images using various types of input, including bounding boxes of the target objects. Hence, the previously extracted YOLO bounding boxes were provided as input to SAM, generating high-quality segmentation masks of the cork strips.

The Figure 3.10 presents a BPMN diagram illustrating the final process for cork strip detection and segmentation. After each image is captured, the strip is first located within the image using the YOLOv11 model. This detection produces a bounding box, which defines the approximate region where the strip is positioned. The resulting bounding box is then passed to the SAM 2 segmentation model, which uses it as a reference point to precisely delineate the strip within the image. After segmentation, the resulting binary mask is stored for use in subsequent stages of the system.

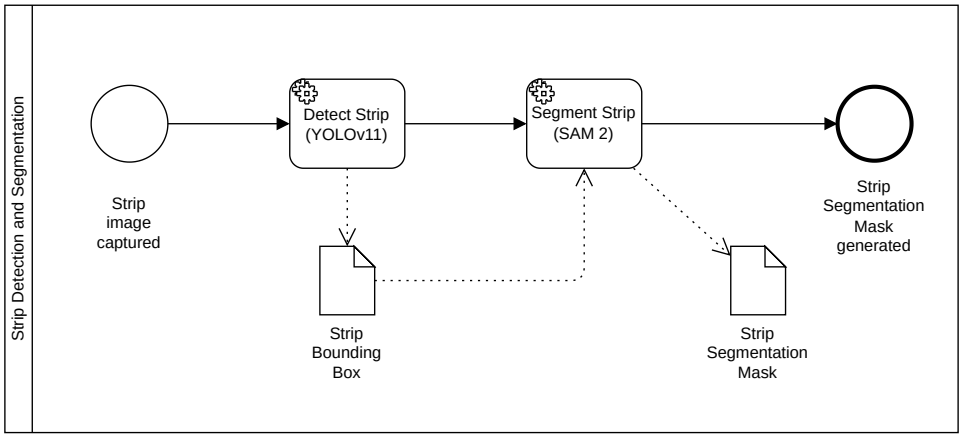


FIGURE 3.10. BPMN diagram of the strip detection and segmentation Process

Since no precise ground-truth annotations were available for these masks, quantitative evaluation was not possible. Nonetheless, visual analysis revealed a substantial improvement in segmentation quality, with the model showing strong adaptability even in irregular or uncommon strip shapes.

Figure 3.11 shows examples of the SAM model applied to two different images. It can be observed that the segmentation of the cork strip is almost perfect, with no noticeable noise or missing regions in the resulting mask.



(A) Example 1 (B) Example 2

FIGURE 3.11. Examples of image segmentation with the SAM model

3.3.2. Holes Detection and Segmentation

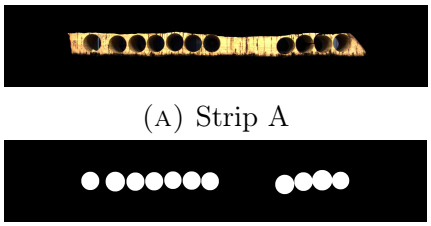
After detecting and segmenting the cork strips, the next step was to detect and segment the holes present in all the strips of the dataset. This step proved essential to understand the underlying pattern of the hole locations and, later, to allow their prediction on images of undrilled strips. To carry out this task, another image processing method was tested, the Hough Circle Transform. According to the documentation of this method, some of the parameters defined in its implementation have a direct influence at the moment of its application. Among the most relevant are:

- `minDist`: minimum distance between the centers of two distinct circles;
- `minRadius` and `maxRadius`: respectively, the minimum and maximum radius of the circles to be detected.
- `param1`: upper threshold used in the edge detection stage (Canny). Controls the sensitivity to edge presence.
- `param2`: threshold applied during the Hough Transform voting process. Defines the minimum number of accumulated votes required for a detection to be accepted as a circle.

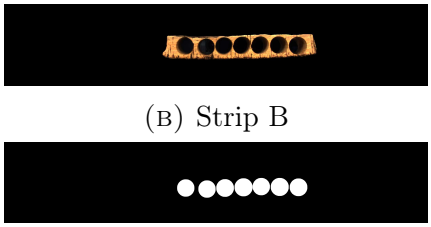
The appropriate values for these parameters were determined empirically, through visual analysis and direct pixel measurements on the images using a photo editor. The choice of the `param1` and `param2` parameters was also decisive for detection quality, since inadequate values can lead either to an excessive number of false positives or to the omission of real holes. Additionally, as recommended in the documentation of the method, an image smoothing process using Gaussian Blur was applied in order to reduce noise and mitigate the incorrect detection of false circles. After multiple iterations, it was verified that this preprocessing, together with the proper selection of detection parameters, is fundamental to obtain consistent results. Since there were no annotations regarding the location of the holes in each strip, it was not possible to quantify performance. However, after visual analysis, the results were globally satisfactory. The main errors observed occurred in cases of partially broken holes: although identified, they interfered with the detection of immediately adjacent holes, occasionally compromising the local precision of the method.

Figure 3.12 shows the result of applying the Hough Circle Transform to two cork strip images. The regions identified as holes are highlighted in white on the segmentation masks, while the rest of the image appears in black. It can be observed that the method

accurately segments the holes, maintaining their positions in the mask consistent with the corresponding hole locations in the original image.



(A) Strip A



(B) Strip B

(C) Strip A hole segmentation mask

(D) Strip B hole segmentation mask

FIGURE 3.12. Examples of hole segmentation

Table 3.2 presents statistics regarding the radii of the detected holes in the dataset, as well as the distances between their centers. The radii show very little variation (standard deviation of only 2.49 pixels). This result is due to the different perspectives in which each strip is photographed, since in reality all holes have the same size, the drill bit that made them has a fixed size. Therefore, it is possible to determine an approximate average hole size of 52 pixels. The distances between holes show greater variability, as some strips have large gaps between holes due to the presence of defective regions. On the other hand, the minimum distance being similar to the average radius suggests that there are holes located very close to each other.

TABLE 3.2. Statistics of the circle radii and distances between centers, in pixels (Px)

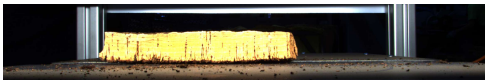
Metric	Radii (px)	Center-to-center distance (px)
min	46	49.820
max	59	2431.695
mean	52.159	519.437
median	52.000	424.001
std	2.494	386.045
P10	49.000	112.219
P25	50.000	219.002
P75	54.000	730.379
P90	55.000	1077.003

3.3.3. Holes Projection onto the undrilled Images

Once the centers of the holes on the strip surfaces were located, it became necessary to determine the corresponding locations on the undrilled surfaces. This step is crucial for the development of the final algorithm, since only in this way is it possible to analyze the visual characteristics of the cork surface at the points where the human operator decided to make the holes. In this context, it became essential to understand whether there

are relevant differences between the images captured before and after drilling, as well as between the left and right views of the same strip. For this purpose, it was necessary to organize the images according to their timestamps and group them coherently: each group of four images corresponds to the same strip, two per view (left and right) and two per moment (before and after drilling). During this analysis, it was verified that the correspondence between views and moments was not always consistent. In some cases, it was observed that, for example, the image of the left view without holes corresponded to the image of the right view with holes and vice versa. This misalignment would make the pipeline malfunction if only the file names were used to group the images. Additionally, it was found that there were differences in orientation, positioning and even deformations of the strips between the moments of capture, which reinforced the need to implement robust alignment and projection procedures to ensure precise correspondence between the locations of the holes and their equivalent positions on the undrilled images.

In Figure 3.13, one of the examples of strips with swapped pairs of images of the same surface can be observed. The columns on the left and right show the images labeled in their names as corresponding to the left and right surfaces, respectively. However, the undrilled image of the left view actually corresponds to the drilled image of the right view and vice versa.



(A) Image A: Image pair in dataset: C; Real image pair: D



(B) Image B: Image pair in dataset: D; Real image pair: C



(C) Image C: Image pair in dataset: A; Real image pair: B



(D) Image D: Image pair in dataset B; Real image pair: A

FIGURE 3.13. Example of strip images with swapped pairs

In Figure 3.14, another example can be observed, in which the pairs of the same surface are swapped and the strip also presents deformation before and after drilling. A loss of curvature can be noticed in the strip on the images corresponding to the moment after drilling, which makes hole projection more difficult. To ensure that the pairs were not swapped, a matching method was implemented. This method is based on the extraction of local descriptors of visual features using the Oriented FAST and Rotated BRIEF (ORB) algorithm, followed by the matching of these descriptors using the Fast Library for Approximate Nearest Neighbors (FLANN) method, configured with an Locality Sensitive Hashing (LSH) index. This method uses the Feature Detection and Description tools available in the OpenCV library, as described in its documentation [36]. For each undrilled image, keypoints are detected and their corresponding descriptors are computed. These are compared with the descriptors of the drilled images from both views. The number of "good" matches is counted.



(A) Image A: Image pair in dataset: C; Real image pair: D



(B) Image B: Image pair in dataset: D; Real image pair: C



(C) Image C: Image pair in dataset: A; Real image pair: B



(D) Image D: Image pair in dataset B; Real image pair: A

FIGURE 3.14. Example of strip images with swapped pairs and with deformations

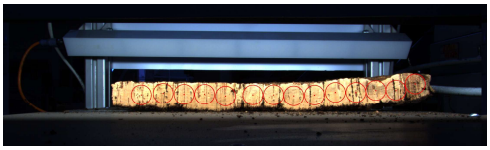
A match is considered "good" if it satisfies Lowe's Ratio Test. According to this test, for each point in the reference image, the two best matches are obtained based on the distance between descriptors. If the ratio between the distance to the best match and the distance to the second-best match is lower than a given threshold, the match is accepted. Otherwise, it is considered ambiguous [37]. The view with the highest number of "good" matches is assumed to be the correct pair. If the best match found for the left view is the image of the right view, the pairing is corrected and recorded.

If there were no deformations between the strips before and after drilling, an alignment of position and orientation would be sufficient, in principle, to project the holes onto the undrilled images. However, these deformations occur quite frequently. Therefore, instead of using the absolute pixel coordinates of the strips, a geometric reference based on the shape of each strip was used, the skeleton. This was obtained by applying the *skeletonize* method from the scikit-image package [38]. The skeleton represents the medial line of the strip, that is, the trajectory running through the center of its structure, regardless of its thickness or curvature. Thus, even if the strip is tilted between captures, slightly curved, or has small scale differences due to its positioning on the table, the skeleton continues to describe the same central axis, allowing correspondence between different "moments" of the same strip [39]. The skeleton is a structure composed of interconnected points that may contain small branches. To ensure that it represents the main axis of the strip, even when slightly curved, the Dijkstra algorithm [40] was used, a method that computes the shortest path between two points in a graph. In this case, each skeleton pixel is a node and each connection between neighboring pixels is an edge with a cost proportional to the distance between them. Thus, the algorithm was used to measure the distances between all pairs of terminal points of the skeleton (the endpoints). Then, the most distant pair was chosen and the shortest path connecting them (without following branches) was calculated. This corresponds to the longest path within the skeleton, resulting in a continuous line that follows the axis of the strip.

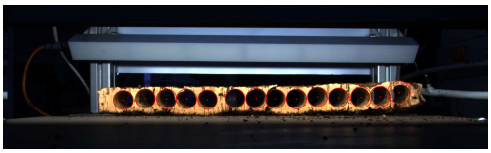
With the main axis identified for both the before-drilling and after-drilling images, each point on the skeleton could be identified by a normalized arc-length parameter representing its relative position along the strip, between 0 (beginning of the skeleton) and

1 (end of the skeleton) [41]. In this way, it was possible to identify points at the same relative position both in the undrilled and in the drilled images. However, during the process of obtaining the skeleton, the order of the points is arbitrary. If not verified, the starting point of one strip could correspond to the end of another, leading to inverted and incorrect projections. Thus, the global orientation of both skeletons was corrected when necessary. For each hole identified in the drilled image, the closest point on the skeleton was found, allowing the arc-length parameter to be obtained. Based on this parameter, the point on the skeleton of the undrilled image with the same value was found, ensuring that the projection occurred at the same relative position along the strip. If the algorithm ended here, all holes would be projected onto the skeleton, which would be incorrect, since there are holes made away from the "center" of the strips. Therefore, the offset of each hole from the skeleton was calculated along a normal vector. This value represents how far and in which direction the hole is displaced from the axis. Thus, the projected point on the undrilled skeleton, at the same relative position, was displaced along its normal in the same direction and distance as the corresponding hole was from the axis of the drilled strip. For visualization of the projection, a fixed radius equal to the mean radius value was used.

Figure 3.15 shows the result of projecting the holes from a drilled image (B) onto its corresponding undrilled image (A). In image (A), the projected holes are marked in red circles. It can be observed that, although both images belong to the same cork strip, their shapes differ slightly due to deformations introduced during the drilling process. Nevertheless, the hole projection method successfully mapped the holes from the drilled image onto the undrilled one, achieving highly consistent positions along the strip.



(A) Undrilled image with projected holes



(B) Image with real holes

FIGURE 3.15. Example of hole projection

3.3.4. Undrilled Strip Alignment

In the context of cork stopper production, when the human operator performs a drill on one surface of a strip, the hole goes completely through the strip. Therefore, the holes coincide on the left and right views of the same strip (when aligned). Thus, to determine whether the model indicates the locations of hole centers to be made in a manner analogous to what a human operator would do, it is not sufficient to analyze only one surface at a time. On one side, at a given position, the strip may have no defects, while on the other side it may have some. Since there are differences between the left and right views of the undrilled images, namely a horizontal flip and positional differences in the image, it was necessary to perform alignment between the undrilled images of both surfaces of the same

strip, so that it would be possible to analyze whether the holes coincide in regions on the opposite surface where drilling should not occur.

For this purpose, the segmented strip masks produced by the SAM model were used, as these define the region occupied by the strip. Because the cameras observe the strip from opposite sides, the first step consisted of mirroring horizontally the image from the right view, so that both had the same orientation. Next, the translation that best aligns the two masks was estimated, using the upper-left corner of the bounding boxes of the strips and attempting to maximize the IOU.

The alignment achieved an average IOU value of 0.7984.

Figure 3.16 presents an example of the alignment between the undrilled surfaces of the same cork strip. In (C), the overlap of the binary masks from both surfaces illustrates the accuracy of the alignment process.



(A) Left surface

(B) Right surface

(C) Overlay of the binary masks of the undrilled surfaces of the same strip

FIGURE 3.16. Example of alignment of undrilled surface images of the same strip

3.3.5. Predicted Holes Filtering

As mentioned previously, the operation of drilling cork stoppers must take into account both surfaces of the strip, in order to avoid defects on either side. In this context, a practical simplification was adopted: it was assumed that all holes are made from the left side of the strip. This choice is consistent with the real operation of the drilling machine, in which the drill bit acts on only one surface. Thus, to ensure the validity of the predicted holes, it was necessary to filter and validate the predictions of the left surface based on the information from the right surface. For this filtering, the alignment method of the undrilled images mentioned in section 3.3.4 was used. Each predicted hole on the left surface was evaluated according to how much of its area was contained within the "good cork" mask projected from the right surface. Holes with at least 80% of their area contained within this valid region were accepted. The others were rejected, as it was assumed that they would pass through defective material on the opposite face.

3.4. Predictive Model Design and Implementation

Before proceeding to the design and implementation of the predictive model, it is necessary to recall how the ground truth was constructed. Initially, the cork strips were detected

and segmented in all images to isolate only the relevant region of interest. Subsequently, the holes were segmented in the drilled images, allowing the extraction of their centers and radii. A correspondence was then established between each undrilled image and its drilled counterpart, ensuring that both represented the same physical strip. Using this correspondence, the holes detected in the drilled images were projected onto the respective undrilled images.

The previously mentioned stages made it possible to generate suitable annotations for the use of supervised learning with the aim of predicting, through undrilled images, the locations of the holes to be made, in a manner analogous to the decision of a human operator. This task was considered a binary segmentation task. The chosen architecture was U-Net, one of the most used structures in image segmentation problems. As illustrated in Figure 3.17, it is characterized by its "U" shaped form, composed of an encoding phase (encoder), responsible for extracting features and progressively reducing the resolution and a decoding phase (decoder), which reconstructs the final mask from those features, restoring the original resolution. This architecture allows the use of skip connections between symmetric layers of the encoder and decoder, which preserve fine spatial information, essential in this case to correctly delineate the regions suitable for drilling [42].

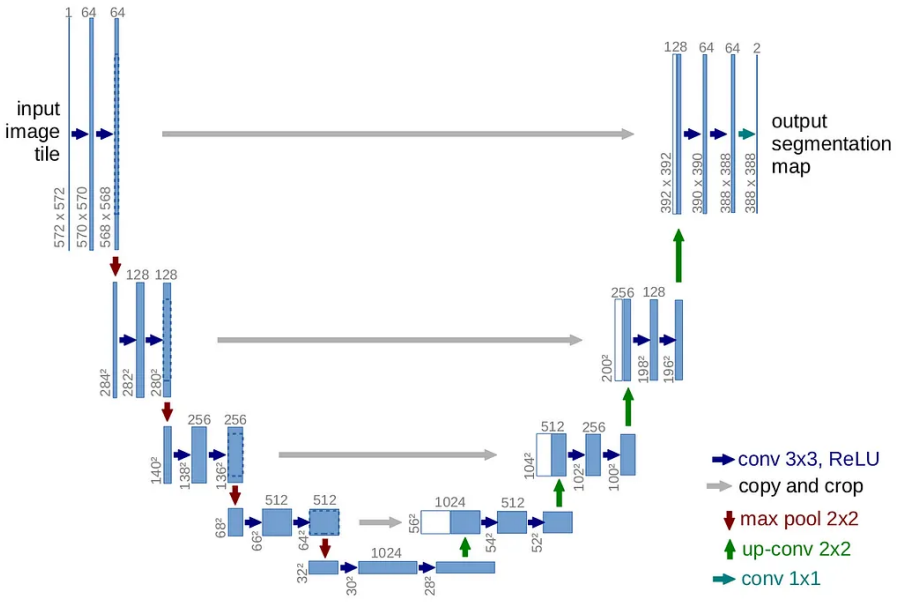


FIGURE 3.17. Classical U-Net architecture diagram [43]

A transfer learning approach [44] was followed, consisting of adapting a pre-trained model to a new task. As backbone, ResNet-34, pre-trained on the ImageNet dataset, was used. This choice is based on the balance between generalization capacity and computational efficiency. The use of pre-trained weights allows the model to start from a base already adapted to generic shapes and textures, accelerating convergence and reducing the need for a large amount of annotated data.

The images used were those already segmented by the SAM model, so that only the cork strip appeared over a black background. They were resized from (2448 × 2048), their original size, to (1216 × 1024) pixels and normalized according to the mean and standard deviation of ImageNet.

During training, the decoder was initialized randomly and trained from scratch, while the encoder was frozen in the first epochs to preserve the previously acquired knowledge. As the model stabilized, the encoder was progressively unfrozen, starting from the deepest layers. This fine-tuning strategy allowed the model to adapt to the specific characteristics of the cork images. Training was performed with small batches due to Graphics Processing Unit (GPU) memory limitations. Optimization was carried out with the AdamW optimizer, widely used in segmentation tasks.

Each pixel in the output image represents the probability of that pixel belonging to the region where a hole should be made (or a "good" region for drilling). The sigmoid activation function normalizes probabilities between 0 and 1. Values close to 1 are considered positive (drilling region) and values close to 0 negative (non-drilling region).

Given the large predominance of background in the images, there is class imbalance: the background (black) occupies most of the image, while the positive regions (holes) represent only a small part of the pixels. Therefore, a composite loss function was used, combining two complementary approaches: Binary Cross Entropy (BCE) [45] and Focal Tversky Loss [46]. BCE was used with additional weighting on the positive class, reinforcing the impact of pixels corresponding to holes. This weighting prevents the model from learning to predict only the background — a common behavior when there is strong class asymmetry. The Focal Tversky Loss complements BCE, being particularly effective in increasing the model's sensitivity to drilling regions, prioritizing the correct detection of pixels belonging to the holes.

During training, evaluation was performed based on two metrics: IOU, calculated with a fixed threshold of 0.5, which measures the overlap between the predicted mask and the reference mask; and F1-Score, determined by sweeping thresholds over the predicted probabilities to identify the optimal balance point between precision and recall. In the context of this problem, precision represents the proportion of pixels predicted by the model as belonging to drilling regions that actually are (i.e., it measures the accuracy of positive predictions), while recall indicates the model's ability to correctly identify all pixels that truly belong to the drilling regions, even those that are difficult to detect. The F1-Score combines both metrics, reflecting the balance between correct predictions and missed detections.

3.5. Software Architecture and Implementation Overview

The implementation of the predictive model and its training pipeline was carried out using the Python programming language, based on the PyTorch deep learning framework. The system was designed following a modular and object-oriented architecture, in order to guarantee reproducibility, scalability and maintainability. Each component

of the system is encapsulated in an independent class or module, allowing clear separation of responsibilities and facilitating updates or the integration of new functionalities without structural modifications. At the data level, a dedicated HolesDataset class was implemented to manage image–mask pairs and ensure consistent preprocessing. This class loads the input images and their corresponding ground-truth masks, applies resizing and normalization and returns each sample as a tensor ready for training. Data augmentation and normalization were implemented using the Albumentations library, while data loading was handled by PyTorch's DataLoader, enabling efficient batching and parallel processing with pinned memory for GPU acceleration. The training process is coordinated by a ModelManager module, which initializes the U-Net model, configures the optimizer, scheduler and loss functions and controls the overall learning process. This module manages the warm-up phase, gradient accumulation and gradient clipping, as well as the periodic saving of model checkpoints based on the best validation F1-Score. Early stopping was also integrated to prevent overfitting and reduce computational time. Learning rate scheduling was implemented through a Cosine Annealing Warm Restarts strategy to allow dynamic adaptation during training. Model validation and metric computation were delegated to a separate Validator class, which performs inference on validation batches and calculates the IOU and F1-Score metrics. This class is also responsible for sweeping through multiple probability thresholds to determine the optimal trade-off between precision and recall, automatically logging the best results.

The overall architecture of the system is represented in Figure A.1, which illustrates the relationships between the main components and the flow of information throughout the training and validation stages. This modular software design ensures flexibility for future experiments, allowing changes in datasets, model architectures, or loss functions with minimal code adaptation.

3.6. Hole Prediction Pipeline

Once the model was trained and validated, inference was performed on the entire test set to predict the regions suitable for drilling. The inference pipeline followed the same preprocessing steps as during training and validation, including resizing and normalization, ensuring consistency across all stages. For each undrilled image, the model produced a probability map representing the likelihood of each pixel belonging to a "good cork" region. These probability maps were thresholded to generate binary masks corresponding to the areas predicted as suitable for drilling. On the left surface, each predicted "good cork" region was filled with circles of fixed radius, producing a set of candidate hole centers. This procedure aimed to simulate the possible drilling positions that the system would recommend based on the inferred material quality. To ensure that these predicted holes respected the quality constraints on both surfaces of the strip, a cross-surface filtering step was applied. For this purpose, the undrilled left and right images were first aligned, as described in Section 3.3.4, so that corresponding regions between the two surfaces could be compared directly. Each predicted hole on the left surface was then evaluated

according to how much of its area was contained within the “good cork” mask projected from the right surface. Holes whose area overlapped at least 80% with valid regions on the opposite side were retained, while the others were discarded. This step ensured that no hole predicted on the left surface would pass through defective material on the right surface, as described in 3.3.5. The remaining holes constitute the final set of drilling centers predicted by the system.

CHAPTER 4

Results and Discussion

The results presented in this chapter are organized into two main stages, reflecting the progressive evaluation of the proposed system. The first stage focuses on assessing the inference capability of the trained model when applied to unseen data, namely the test set composed of undrilled images. In this step, the objective was to evaluate how effectively the model identifies regions of cork suitable for drilling, the so-called "good cork" areas.

The second stage addresses the geometric evaluation, where the predicted drilling regions are converted into discrete hole centers to allow direct comparison with real holes extracted from the corresponding drilled images. This step provides a more application-oriented assessment, as it evaluates not only whether the model identifies suitable drilling zones but also whether the predicted hole locations are spatially consistent with those produced by human operators.

This two-level analysis, first at the segmentation level and then at the geometric level, enables a comprehensive understanding of the model's performance, from the detection of suitable material regions to the precise prediction of drilling coordinates.

4.1. Evaluation Metrics

To objectively assess the performance of the proposed system, two complementary levels of evaluation were conducted, each addressing a different stage of the processing pipeline.

At the first level, corresponding to the segmentation stage, the objective was to measure how effectively the model identifies regions of cork suitable for drilling. For this purpose, the probability maps produced during inference were thresholded to generate binary masks representing "good cork" areas. The resulting masks were compared with the ground-truth reference masks obtained from the drilled images. The evaluation was performed in terms of pixel-level metrics, namely Intersection over Union (IoU), Precision, Recall and F1-Score. These indicators provide a quantitative understanding of how well the model distinguishes between suitable and unsuitable regions of cork.

At the second level, corresponding to the prediction of drilling centers, a geometric comparison was performed between the predicted and real hole coordinates. Each image in the test set contains a set of reference points (ground truth) and a corresponding set of predicted centers. Since the order and number of predicted and reference points may differ, the Hungarian algorithm (also known as the linear sum assignment method) was employed to optimally match predictions to ground-truth points based on their pairwise Euclidean distances. A tolerance threshold of 60 px (approximately equal to the mean hole radius) was adopted to define valid matches. Only pairs within this distance were

considered correct detections. From these correspondences, the classical metrics Precision, Recall and F1-Score were computed, while additional indicators such as the mean and 95th percentile distances were used to characterize the spatial accuracy of the predictions. Furthermore, differences between the predicted and ground-truth hole counts per image were analyzed to assess the overall consistency of the model's outputs across the dataset.

4.1.1. Segmentation Level Assessment Metrics

To quantitatively assess the quality of the segmentation masks produced by the model, a set of standard pixel-level metrics was used. These metrics compare the predicted regions with the corresponding ground-truth annotations, providing an objective measure of similarity and overall performance.

Let $\hat{Y}, Y \in \{0,1\}^{H \times W}$ denote, respectively, the predicted binary mask obtained after thresholding the probability map and the ground-truth mask that defines the reference regions suitable for drilling. Each element $\hat{Y}(u,v)$ or $Y(u,v)$ thus represents a binary label associated with pixel coordinates (u,v).

From these definitions, we can compute the fundamental components of the confusion matrix:

$$TP = \sum_{u,v} [\hat{Y}(u,v) = 1 \wedge Y(u,v) = 1], \tag{4.1}$$

$$FP = \sum_{u,v} [\hat{Y}(u,v) = 1 \wedge Y(u,v) = 0], \tag{4.2}$$

$$FN = \sum_{u,v} [\hat{Y}(u,v) = 0 \wedge Y(u,v) = 1]. \tag{4.3}$$

Here, TP (True Positives) counts the correctly identified pixels within the valid drilling regions, FP (False Positives) represents pixels incorrectly predicted as valid and FN (False Negatives) corresponds to real positive pixels that were missed by the model.

Based on these counts, the pixel-level metrics are computed as:

$$\text{Precision} = \frac{TP}{TP + FP} \tag{4.4}$$

$$\text{Recall} = \frac{TP}{TP + FN} \tag{4.5}$$

$$\text{F1-score} = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \tag{4.6}$$

$$\text{IoU} = \frac{TP}{TP + FP + FN} \tag{4.7}$$

Here:

- Precision quantifies the proportion of pixels predicted as "good cork" that truly belong to the reference region;
- Recall measures the proportion of real positive pixels correctly identified;
- F1-score provides a balanced indicator of overall segmentation performance;

- IoU (Intersection over Union) captures the overlap between predicted and ground-truth masks, being a stricter indicator of segmentation quality.

4.1.2. Geometric Level Assessment Metrics

While segmentation metrics assess the pixel-wise quality of the predicted regions, they do not directly measure how accurately the model predicts the geometric position of each drilling point. Therefore, the final geometric evaluation was conducted to compare the predicted and ground truth hole centers. This analysis focuses on spatial correspondence, quantifying both the positional accuracy and the number of correctly matched holes through distance-based matching.

To evaluate the geometric accuracy of the predicted hole locations, let $\mathcal{P} = \{p_i = (x_i, y_i)\}_{i=1}^{N_P}$ denote the set of predicted centers and $\mathcal{G} = \{g_j = (x_j, y_j, r_j)\}_{j=1}^{N_G}$ the set of ground truth centers, where r_j corresponds to the true radius of each hole.

The pairwise Euclidean distance matrix between ground-truth and predicted centers is computed as:

$$D_{j,i} = \|g_j - p_i\|_2 = \sqrt{(x_j - x_i)^2 + (y_j - y_i)^2}.$$

The optimal correspondence between predictions and ground truth points is obtained using the Hungarian algorithm, which solves the following minimum cost assignment problem:

$$\min_{M} \sum_{(j,i)\in M} D_{j,i},$$

subject to one-to-one matching constraints that ensure each prediction and ground-truth center can only be paired once.

A tolerance threshold $\tau = 60\,\text{px}$ is applied to discard mismatches; only assignments satisfying $D_{j,i} \leq \tau$ are considered valid, while distances exceeding this limit are assigned an infinite cost.

From the resulting set of valid matches M_{valid}, the evaluation metrics are defined as:

$$TP = |M_{\text{valid}}|, \tag{4.8}$$

$$FP = N_P - TP, \tag{4.9}$$

$$FN = N_G - TP. \tag{4.10}$$

Based on these counts, the standard evaluation metrics are computed as follows:

$$\text{Precision} = \frac{TP}{TP + FP} \tag{4.11}$$

$$\text{Recall} = \frac{TP}{TP + FN} \tag{4.12}$$

$$\text{F1-score} = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \tag{4.13}$$

These metrics respectively measure:

- Precision — the proportion of predicted centers that correspond to real holes;
- Recall — the proportion of real holes correctly detected;

- F1-score — the harmonic mean of precision and recall, penalizing large imbalances between the two.

4.2. Segmentation Level Evaluation

For the evaluation of the "good cork" regions, inference was performed on the test set, composed of 394 images. The inference pipeline followed the same transformations used during validation (resizing and normalization), ensuring consistency between training, validation and testing. From each image, a probability map was obtained. To extract the region corresponding to areas suitable for drilling, it was necessary to define a threshold to classify a pixel as positive. A sweep over the probability map of the test set was performed to characterize the trade-off between precision and recall, as well as to identify the point of highest F1-Score. It is important to clarify that this evaluation is not the final one but rather complementary to it. The resulting metrics were:

- IOU @ 0.5 = 0.5505
- Precision = 0.652
- Recall = 0.845
- F1-Score = 0.736

In the context of this problem, precision quantifies the proportion of pixels marked as "good cork" that are, in fact, within the positive reference region. Recall measures the fraction of real positive pixels that were correctly identified as such.

By analyzing the results, it was concluded that the model is capable of detecting most of the positive regions (due to the high recall). Thus, the probability of not detecting areas that are indeed good for drilling is low. On the other hand, the more moderate precision indicates some error in the regions marked as "good," suggesting the presence of false positives. The high F1-Score reflects a reasonable balance between precision and recall, suggesting that the model generalizes satisfactorily. Furthermore, a tendency toward more false positives rather than more false negatives benefits the final result. The analysis of real cork stopper production processes shows that after the drilling stage, there is a quality characterization of the produced stoppers. When a stopper does not meet the necessary quality standards to be used, it is reused for the production of cork derivatives [3]. Therefore, it can be concluded that this model promotes a greater utilization of cork. If any stopper is drilled but does not possess the required quality, it will be detected in subsequent stages.

To allow a direct comparison between real holes and predicted ones, a method was implemented that, for each image, filled its "good cork" mask with circles of fixed radius. This made it possible to visually analyze how the predicted holes covered the regions considered suitable for drilling, helping to assess the geometric plausibility of the method.

Figure 4.1 presents an example of the inference result obtained on the test set. The figure includes two views. In (A), an overlay of the predicted "good cork" mask (outlined in red) is shown together with the ground-truth holes (marked in green). The "good

cork" mask was obtained by applying a threshold to the probability map produced by the model. It can be observed that not the entire strip is covered by the mask, which indicates the presence of visually defective areas. In view (B), the same mask is filled with circles of fixed radius, representing potential drilling holes. This visualization is used to evaluate how accurately the mask was predicted, enabling a direct comparison between the predicted and real holes.



(A) Undrilled image with real holes marked in green and drilling regions outlined in red



(B) Drilling regions filled with fixed-radius circles

FIGURE 4.1. Test set inference sample

For quantitative performance measurement, the predicted centers were compared with the real centers. Specifically, a Euclidean distance matrix was built between all pairs and the Hungarian algorithm [47] (global optimal assignment) was applied on a one-to-one basis but with a tolerance. Pairs whose distance exceeded this tolerance were not considered corresponding. For a tolerance equal to the average radius value, 52 pixels, the following results were obtained:

- Precision: 0.772
- Recall: 0.918
- F1-Score: 0.839
- Mean-Distance: 27.15 pixels

The high recall shows that most real holes coincided with the predicted ones (few false negatives). On the other hand, the moderate precision indicates some false positives. The distances among true positives suggest satisfactory geometric alignment between the predicted and real hole centers, with controlled outliers.

It is important to mention that sometimes the number of holes extracted from different surfaces of the same strip varies, which should not happen. This occurs because, in some cases (e.g., broken holes), the model may detect only one where there are more. For example, applying the same model but only to cases where the number of predicted and ground-truth holes matches produces more promising metrics:

- Precision: 0.892
- Recall: 0.892
- F1-Score: 0.892
- Mean-Distance: 25.61 pixels

In this case, the performance clearly improves, with an F1-Score of 0.892 and lower distances.

4.3. Geometric Level Evaluation

The geometric evaluation of the model was conducted on the test set, considering only the undrilled images. The results are as follows:

Considering the metrics presented in Table 4.1, the precision of 0.861 showed that most of the holes predicted by the model corresponded to real holes, indicating a low occurrence of false positives (holes predicted in regions where no real holes exist — i.e., in "bad cork" areas). On the other hand, the recall of 0.911 indicated that the model was able to identify practically all existing holes, with only a small portion of real holes not detected (false negatives). The F1-Score value of 0.886 demonstrated a balance between these two dimensions, showing that the model achieved a good compromise between avoiding incorrect predictions and not omitting relevant holes.

TABLE 4.1. Aggregated results on the test set (Hungarian matching; tolerance = 60 px)

Metric	Value
TP	1461
FP	235
FN	143
Precision	0.861
Recall	0.911
F1-score	0.886

Complementarily, the confusion matrix illustrated in Figure 4.2 graphically summarizes the relationship between correct and incorrect predictions. The predominance of the quadrant corresponding to true positives confirms the high coherence between the predicted and real patterns.

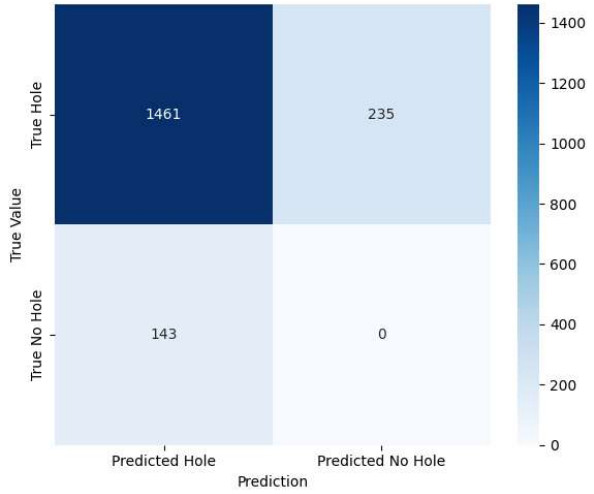


FIGURE 4.2. Confusion Matrix - Final Evaluation

The combined analysis of these metrics allows concluding that the model presents a consistent and robust performance in reproducing the drilling patterns observed in the strips.

In addition to the spatial correspondence between predicted and real holes, the model's consistency was also analyzed in terms of the total number of predicted holes per image.

Table 4.2 presents a quantitative analysis of the predicted and real hole counts in images where both real annotations and model predictions exist. It can be observed that the model predicts, on average, 8.61 holes per image, a value very close to the real mean of 8.14 holes, corresponding to an average deviation of only 0.47 holes. The mean absolute deviation, below one hole, demonstrates the high stability of the model in terms of the number of holes predicted per strip.

TABLE 4.2. Summary of hole counts — intersection (images present in GT and Pred)

Metric	Value
No. of images	197
Mean holes (GT)	8.14
Mean holes (Pred)	8.61
Mean deviation (Pred−GT)	+0.47
Mean absolute deviation	0.95
Count: equal	83
Count: over (Pred > GT)	82
Count: under (Pred < GT)	32

Figure 4.3 shows the distribution of the difference between predicted and real hole counts, 83 images have exactly the same number of predicted and real holes, 82 show a slight overprediction and 32 show a small underprediction. These results indicate that the model maintains a balanced behavior, with a minimal tendency to produce slightly more holes than those generated by the human operator. This is acceptable in the production context, since it does not lead to cork waste and redundant quality analysis methods exist for the produced stoppers.

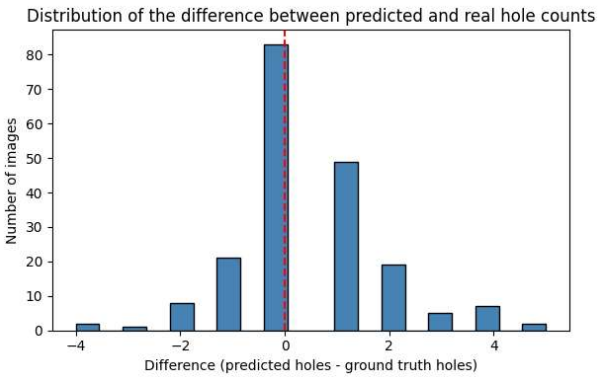


FIGURE 4.3. Distribution of the difference between predicted and real hole counts

4.4. Qualitative Results

In addition to the quantitative metrics presented earlier, a qualitative analysis of the prediction results was carried out, with the goal of visually assessing the spatial coherence of the predicted holes and their correspondence with the drilling pattern observed in the real images.

The following figures present sample results of the predictions, annotated as follows:

- Green circles represent ground-truth holes (GT);
- blue circles correspond to true positives (TP);
- yellow lines connect each true positive to its corresponding ground-truth hole, as determined by the Hungarian matching;
- red crosses indicate false positives (FP).

If a green circle (GT) isn't connected to a blue circle, it is a (FN).

Figure 4.4 shows a strip where the hole prediction was particularly accurate. It can be observed that almost all predicted holes, except one, have an associated ground-truth counterpart. Moreover, the yellow lines connecting each prediction to its corresponding ground-truth hole are short, indicating a strong spatial alignment between predicted and real positions, a characteristic strength of the model.

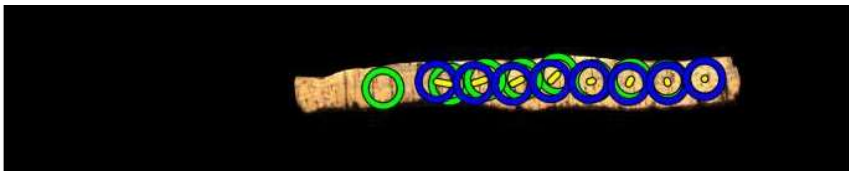


FIGURE 4.4. Example of hole prediction (1)

Figure 4.5 shows a strip where two (FN) are present. These holes are located in a region whose texture visually resembles areas typically associated with defects. When analyzing the dataset, a few annotations of this type can be found, which slightly deviate from the usual labeling pattern. This may be related to the repetitive nature of the manual annotation process and the potential for occasional operator error.



FIGURE 4.5. Example of hole prediction (2)

Figure 4.6 illustrates an uncommon error produced by the model. Although it generally tends to generate more false positives (FP) than false negatives (FN), it rarely places holes in regions located near visible defects. Despite this isolated error, the remaining predictions are correct and well aligned with the ground truth.

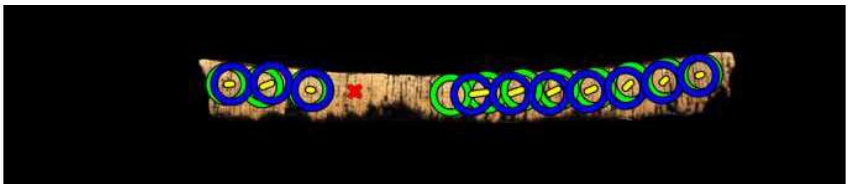


FIGURE 4.6. Example of hole prediction (3)

Figure 4.7 presents a case in which the model performs particularly well. The strip has a stable and regular shape, with no significant deformations and the ground-truth holes almost completely cover its length. This characteristic may reflect the high quality of the cork material in this specific sample.

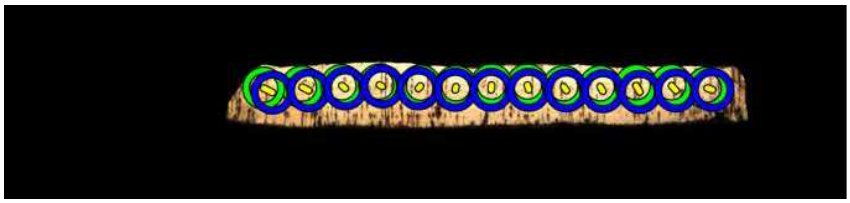


FIGURE 4.7. Example of hole prediction (4)

Figure 4.8 allows drawing some recurring conclusions about the model's behavior. First, the central fractured region of the strip has no predicted holes, which is a common outcome in this model. Second, a false negative (FN) can be observed on the right side of the strip. This may be related to the shadow visible in the upper part of the image, which could have led the model to interpret that area as defective. On the other hand, the false positive (FP) located on the left side illustrates a case where the predicted hole is slightly farther from the ground-truth hole than the defined tolerance of 60 px used for the matching. Nevertheless, it is possible to observe that the general drilling pattern of the ground truth is largely replicated by the model.



FIGURE 4.8. Example of hole prediction (5)

Figure 4.9 shows another case of almost perfect overlap between the ground truth and the predicted holes. It can be observed that the distance between the predicted and ground truth centers is very small, reinforcing the model's ability to accurately predict drilling positions in strips with regular shapes and minimal deformations.

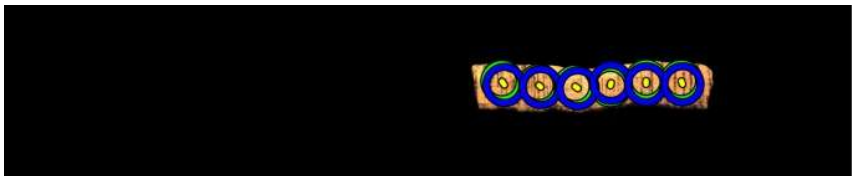


FIGURE 4.9. Example of hole prediction (6)

Figure 4.10 illustrates a particularly relevant case for analysis. The strip appears to contain a large surface area with visible defects; however, the ground truth still includes holes within those regions. In such situations, the model tends to behave more conservatively, avoiding predictions in areas that visually exhibit strong texture irregularities. It can be observed that the predicted holes are mostly located in regions where the cork texture appears normal and homogeneous.



FIGURE 4.10. Example of hole prediction (7)

Figure 4.11 shows another case where the model avoids making predictions in regions located near major variations in the cork texture. In the other hand, in areas where the texture appears normal and homogeneous, the model successfully places its predictions in accurate positions.

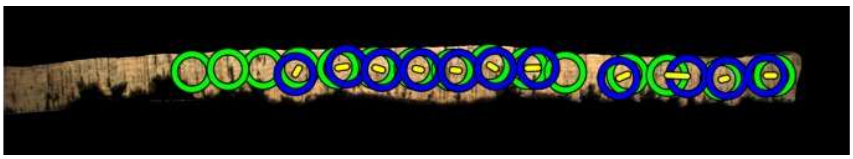


FIGURE 4.11. Example of hole prediction (8)

Appendix B provides an end-to-end example of the system applied to a single cork strip, covering annotation construction, hole prediction and filtering.

4.5. Discussion

The results obtained demonstrate that the proposed model is capable of predicting, with high accuracy, the locations of the holes to be made in undrilled cork strips. With an F1-Score of 0.8855, a Precision of 0.861 and a Recall of 0.911, the model presents a consistent balance between the ability to correctly identify drilling zones and the limitation of false positives. These values, together with the visual analysis of the predictions, indicate that the system is able to effectively reproduce the drilling pattern generated by the human operator. The slight difference between Precision and Recall suggests that the model tends to predict slightly more holes than those that actually exist. This fact is

confirmed by the analysis of the number of predicted versus real holes, where a positive mean deviation of +0.47 holes per image was observed. The qualitative analysis reinforces these conclusions. In strips with homogeneous texture and consistent lighting, an almost perfect overlap between predicted and real holes was observed (Figures 4.4, 4.7 and 4.9) showing that the model not only learned to identify the "good cork" regions but also captured the geometric regularity of the drilling process. In these cases, the distances between predicted and real centers were minimal and the alignment of the holes along the strip closely followed the expected pattern of a human operator, confirming the success of the supervised learning–based approach.

Figures 4.5 and 4.8 illustrate situations where false negatives occurred, typically in regions affected by shadows or local surface fractures. These visual irregularities can lead the model to classify such areas as defective, even when the ground truth indicates otherwise. Figures 4.10 and 4.11 further show that the model tends to avoid predictions in areas exhibiting strong texture variations or large visual defects. In these situations, the predicted holes are concentrated in portions of the strip with more uniform texture, which aligns with the expected human decision to avoid drilling through defective material. Overall, the analysis confirms that the proposed model not only generalizes well to unseen data but also mimics the human reasoning underlying the drilling process. Errors mainly occurred in areas with severe surface irregularities, segmentation artifacts, or inconsistent illumination. These limitations are consistent with the natural variability of cork and the subjective nature of human annotation. Nevertheless, the model exhibited robustness to small deformations and scale variations, largely due to the use of the skeleton as a geometric reference for hole projection.

In summary, the results demonstrate that the proposed architecture is well suited to the task, presenting strong quantitative performance and qualitative behavior that faithfully replicates the human operator's drilling patterns.

CHAPTER 5

Conclusions and Future Work

5.1. Conclusions

The work developed in this dissertation addressed the challenge of automatically predicting drilling centers on cork strips prior to the drilling stage, aiming to replicate, as faithfully as possible, the decision-making process of human operators. In the traditional industrial setting, this task is still carried out manually: an operator visually inspects both surfaces of each cork strip and decides where to place the holes, taking into account texture, defects and spatial balance. This activity, while fundamental to ensuring product quality, is time-consuming, repetitive and physically demanding and it inherently introduces variability associated with subjective human judgment.

The motivation for automating this process arises from three main needs in the cork industry:

- increasing production capacity, by reducing the dependency on manual operations;
- improving consistency and precision, ensuring uniformity in drilling decisions; and
- enhancing operator safety, by minimizing repetitive strain and exposure to potentially hazardous industrial environments.

In this context, the work aimed to design a system capable of automate the decision of where to drill, ensuring that the resulting patterns mai+ntained the same quality and logic as those achieved manually.

To achieve this goal, a complete computer vision pipeline was developed, encompassing the main stages of perception and reasoning involved in the task. First, the cork strip is automatically detected and segmented in each image, isolating it from the background to define the region of interest. Then, both surfaces (left and right) are aligned so that each point on one side corresponds precisely to its counterpart on the other. This alignment ensures that the model evaluates each potential drilling region in an integrated way, considering the quality of both surfaces. Subsequently, the ground truth was built by detecting holes in drilled strips and projecting their positions onto the corresponding undrilled versions. This process provided accurate reference coordinates for the supervised learning stage.

The proposed predictive model, based on a U-Net architecture with a ResNet-34 encoder pre-trained on ImageNet, was trained to identify regions suitable for drilling, learning to associate textural patterns and geometric configurations with successful drilling

locations. The transfer learning strategy proved particularly valuable, enabling the model to converge efficiently even with a relatively limited dataset and to generalize effectively across strips with different visual and structural characteristics.

From an experimental perspective, the system achieved strong and consistent results. In the segmentation level evaluation, the model obtained an IoU of 0.5505, a Precision of 0.652 and a Recall of 0.845, demonstrating its ability to detect most of the "good cork" regions while maintaining an acceptable level of false positives. At the geometric level, when comparing predicted drilling centers with the ground truth, the model achieved an F1-Score of 0.8855, a Precision of 0.861 and a Recall of 0.911, confirming a high degree of spatial alignment between predicted and real holes. The mean distance between matched holes (27.15 px) and the 95th percentile of distances (48.84 px) were within the acceptable tolerance margin (60 px), further reinforcing the geometric accuracy of the predictions.

The qualitative analysis complemented these findings by visually illustrating the behavior of the model under different conditions. In most cases, a near perfect overlap was observed between the predicted and ground-truth holes, especially in strips with stable geometry and uniform texture. The model shows a slight tendency to overpredict: it occasionally proposes holes in regions that were not drilled by the human operator, particularly near the edges of the strip or in thinner portions where the available space makes the decision more ambiguous. From an industrial standpoint, this tendency toward mild overprediction is acceptable. Slightly increasing the number of suggested drilling positions does not imply material waste, since each stopper is still subject to downstream quality inspection and non-compliant stoppers are diverted to derivative uses. Overall, these observations confirm that the model not only captures the visual cues that human operators rely on, but also reproduces much of their spatial reasoning, while erring on the side of proposing additional viable drilling locations rather than missing exploitable cork.

The conclusions drawn from these results provide clear answers to the research questions formulated in this work:

- RQ1 – Can computer vision models effectively reproduce the decision-making process of human operators in cork stopper drilling, achieving drilling patterns comparable to those obtained through manual intervention?
  - Yes. The proposed model successfully reproduced the spatial logic and decision-making process typically performed by human operators. The predicted hole distributions exhibited geometric regularity, spacing and alignment consistent with manual drilling patterns, validating the feasibility of an automated solution.
- RQ2 – How do data preparation and pre-training strategies affect the performance and generalization ability of computer vision models in predicting drilling positions on cork strips?

  - The experiments confirmed that the data preparation pipeline, particularly the alignment of surfaces and the projection of drilling points, plays a crucial role in achieving consistent performance. Additionally, transfer learning using a ResNet-34 backbone significantly improved generalization and stability, allowing the model to learn complex textural features of cork with fewer data and in fewer epochs.
- RQ3 – Is it possible to adapt a segmentation model to accurately predict drilling patterns on cork strips, ensuring robustness and consistency across different visual and structural conditions?
  - The U-Net segmentation model proved to be effective, offering a favorable trade-off between accuracy, interpretability and computational efficiency. Its encoder–decoder design with skip connections preserved fine spatial details while maintaining contextual understanding, which was essential for identifying suitable drilling regions even in the presence of visual noise or geometric deformation.

In summary, the developed system demonstrates that computer vision and deep learning techniques can successfully automate the drilling decision process, achieving results that parallel human performance in both accuracy and reasoning logic. Beyond improving precision and consistency, the system contributes to increased production capacity and operator safety, aligning with the current digital transformation trends in the indutry.

5.2. Future Work

Although the proposed system achieved promising results, several limitations were identified throughout the research, both in the data preparation phase and in the predictive modeling stage. These limitations define clear directions for future improvements.

One of the main challenges encountered was the construction of the ground truth annotations, particularly the process of projecting the holes from drilled images onto their corresponding undrilled images. This step required precise geometric alignment between pairs of images captured under slightly different conditions, which occasionally introduced small spatial inconsistencies. Future work should explore more robust image registration techniques, to ensure more accurate projection of hole coordinates across views.

Another limitation was related to the segmentation of the holes used to generate the ground truth data. The Hough Circle Transform, while effective in most cases, presented difficulties in handling broken or irregularly shaped holes, leading to occasional detection gaps or inaccurate radii estimation. Future research could investigate the use of deep learning based segmentation models (e.g., U-Net, Mask R-CNN, or SAM based refinements) to obtain more precise and complete masks of the drilled holes, especially in challenging visual conditions.

From an operational standpoint, a natural continuation of this work is the integration of the predictive module into an actual industrial drilling machinery, enabling real-time decision-making and adaptive learning through operator feedback. This would transform

the system from a proof of concept into a deployable industrial tool, directly contributing to the automation, safety and productivity of the cork stopper production process.

References

[1] "Cork oak forest - the origin of cork," Amorim Cork, Accessed: Oct. 19, 2025. [Online]. Available: https://www.amorim.com/en/cork/cork-oak-forest/.

[2] W. M. P. Office, "Cork screwed? environmental and economic impacts of the cork stoppers market," WWF/MEDPO, Rome, Italy, Tech. Rep., 2006, Available: https://wwfeu.awsassets.panda.org/downloads/cork_rev12_print.pdf.

[3] Amorim Cork, *Raw material and production process*, Accessed: 18 October 2025, 2025. [Online]. Available: https://www.amorimcork.com/en/cork-and-wine/raw-material-and-production-process/.

[4] APCOR, *The cork sector - characterization*, Accessed: 18 October 2025, 2023. [Online]. Available: https://apcor.pt/uploads/homepage/destaques/3_O%20Setor%20da%20Cortica_Caracterizacao_EN.pdf.

[5] J. Cunha et al., "An application of machine vision for automating the manufacture of cork stoppers," in *Proceedings of the 1994 IEEE International Symposium on Industrial Electronics (ISIE'94)*, 1994, pp. 374–379. DOI: 10.1109/ISIE.1994.333087.

[6] K. Peffers et al., "The design science research process: A model for producing and presenting information systems research," *arXiv*, Feb. 2020, Preprint.

[7] B. Kitchenham, "Procedures for performing systematic reviews," Keele University and National ICT Australia Ltd., Tech. Rep., 2004.

[8] J. Gonzalez-Adrados et al., "Quality grading of cork planks with classification models based on defect characterisation," *Holz als Roh- und Werkstoff*, vol. 58, pp. 39–45, 2000. DOI: 10.1007/s001070050383.

[9] A. Salgueiro et al., "Neural network classifier and robotic manipulation for an autonomous industrial cork feeder," in *Progress in Artificial Intelligence: Proceedings of the 20th EPIA Conference on Artificial Intelligence*, 2021, pp. 433–444. DOI: 10.1007/978-3-030-86230-5_34.

[10] Q. Yunlong et al., "Comparison of object detection algorithm using yolo vs faster r-cnn: A systematic literature review," in *Proceedings of the 6th International Conference on Big Data Technologies*, 2023, pp. 419–424. DOI: 10.1145/3627377.3627443.

[11] G. Bradski, A. Kaehler, and the OpenCV Team, *Opencv: Open source computer vision library*, version 4.x, 2025. Accessed: Oct. 19, 2025. [Online]. Available: https://opencv.org/.

[12] J. Cunha et al., "Computer vision and robotic manipulation for automated feeding of cork drillers," *Materials and Design*, vol. 82, pp. 290–296, 2015. DOI: 10.1016/j.matdes.2015.05.037.

[13] F. Alves, "Segmentation and manipulation of cork strips in bulk," in *2021 IEEE International Conference on Autonomous Robot Systems and Competitions (ICARSC 2021)*, 2021, pp. 149–154. DOI: 10.1109/ICARSC52212.2021.9429769.

[14] B. Paniagua et al., "3d textural mapping and soft-computing applied to cork quality inspection," *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, vol. 5358, no. Part 1, pp. 743–752, 2008. DOI: 10.1007/978-3-540-89639-5_71.

[15] J. Lima et al., "A modular approach to real-time cork classification using image processing," in *IEEE International Conference on Emerging Technologies and Factory Automation (ETFA)*, 2005, pp. 361–368.

[16] B. Paniagua et al., "A solution for evaluating the stopper quality in the cork industry," in *ICINCO 2006 - 3rd International Conference on Informatics in Control, Automation and Robotics*, 2006, pp. 334–339.

[17] P. Radeva et al., "Bayesian classification for inspection of industrial products," *Lecture Notes in Artificial Intelligence (Subseries of Lecture Notes in Computer Science)*, vol. 2504, pp. 399–407, 2002. DOI: 10.1007/3-540-36079-4_35.

[18] B. Paniagua et al., "Can fuzzy logic improve the cork quality classification? a comparative study," in *Proceedings of the Mediterranean Electrotechnical Conference (MELECON)*, 2008, pp. 805–810. DOI: 10.1109/MELCON.2008.4618534.

[19] M. A. Vega-Rodríguez et al., "Cork stopper classification using fpgas and digital image processing techniques," in *Proceedings - Euromicro Symposium on Digital Systems Design: Architectures, Methods and Tools (DSD 2001)*, 2001, pp. 270–275. DOI: 10.1109/DSD.2001.952295.

[20] B. Paniagua et al., "Finding the best classifier for evaluating cork quality in an industrial environment," *Lecture Notes in Electrical Engineering*, vol. 15, pp. 183–194, 2008. DOI: 10.1007/978-3-540-79142-3_15.

[21] B. Paniagua et al., "Image processing and neuro-fuzzy computing for cork quality classification," in *IEEE International Conference on Industrial Informatics (INDIN)*, vol. 2, 2007, pp. 657–661. DOI: 10.1109/INDIN.2007.4384851.

[22] B. Paniagua et al., "Improving the industrial classification of cork stoppers by using image processing and neuro-fuzzy computing," *Journal of Intelligent Manufacturing*, vol. 21, no. 6, pp. 745–760, 2010. DOI: 10.1007/s10845-009-0251-4.

[23] A. Costa et al., "Decision rules for computer-vision quality classification of wine natural cork stoppers," *American Journal of Enology and Viticulture*, vol. 57, no. 2, pp. 210–219, 2006. DOI: 10.5344/ajev.2006.57.2.210.

[24] A. Costa et al., "Quality characterization of wine cork stoppers using computer vision," *Journal International des Sciences de la Vigne et du Vin*, vol. 39, no. 4, pp. 209–218, 2005. DOI: 10.20870/oeno-one.2005.39.4.887.

[25] L. Qu, G. Chen, K. Liu, and X. Zhang, "A study on the improvement of yolov5 and the quality detection method for cork discs," *Photonics*, vol. 11, no. 9, p. 825, 2024. DOI: 10.3390/photonics11090825. [Online]. Available: https://www.mdpi.com/2304-6732/11/9/825.

[26] W. Zhou, Y. Li, L. Liu, H. Wang, and M. You, "Cork classification based on multi-scale faster-r-cnn with machine vision," *Measurement*, vol. 217, p. 113 089, 2023, Available: https://www.sciencedirect.com/science/article/pii/S026322412300653X. DOI: 10.1016/j.measurement.2023.113089.

[27] A. Guimarães, M. Valério, B. Fidalgo, R. Salas-Gonzalez, C. Pereira, and M. Mendes, "Cork oak production estimation using a mask r-cnn," *Energies*, vol. 15, no. 24, p. 9593, 2022. DOI: 10.3390/en15249593. [Online]. Available: https://doi.org/10.3390/en15249593.

[28] A. M. Mezher and A. E. Marble, "A novel strategy for improving robustness in computer vision manufacturing defect detection," *arXiv preprint arXiv:2305.09407*, 2023. DOI: 10.48550/arXiv.2305.09407. [Online]. Available: https://arxiv.org/abs/2305.09407.

[29] A. Vijayakumar, S. Vairavasundaram, J. A. S. Koilraj, M. Rajappa, K. Kotecha, and A. Kulkarni, "Real-time visual intelligence for defect detection in pharmaceutical packaging," *Scientific Reports*, vol. 14, p. 18 811, 2024. DOI: 10.1038/s41598-024-69701-z. [Online]. Available: https://www.nature.com/articles/s41598-024-69701-z.

[30] T. Luo, "Pre-training for industrial defect detection," in *Proceedings of the 2023 5th International Academic Exchange Conference on Science and Technology Innovation (IAECST)*, Guangzhou, China: IEEE, 2023, pp. 1475–1480. DOI: 10.1109/IAECST60924.2023.10503045. [Online]. Available: https://ieeexplore.ieee.org/document/10503045.

[31] R. Augustauskas, A. Lipnickas, and T. Surgailis, "Segmentation of drilled holes in textured wooden furniture panels using deep neural network," *Sensors*, vol. 21, no. 11, p. 3633, 2021. DOI: 10.3390/s21113633. [Online]. Available: https://www.mdpi.com/1424-8220/21/11/3633.

[32] G. Jocher and J. Qiu, *Ultralytics yolo11 documentation*, https://docs.ultralytics.com/models/yolo11/, Accessed: Oct. 20, 2025, 2024.

[33] W. Kentaro et al., *Labelme — image polygonal annotation tool*, https://labelme.io/, Accessed: Oct. 20, 2025, 2024.

[34] G. Jocher and J. Qiu, *Yolo performance metrics — ultralytics yolo docs*, https://docs.ultralytics.com/guides/yolo-performance-metrics/, Accessed: Oct. 20, 2025, 2024.

[35] G. Jocher and J. Qiu, *Sam 2: Segment anything model 2 — ultralytics yolo docs*, https://docs.ultralytics.com/models/sam-2/, Accessed: Oct. 20, 2025, 2024.

[36] A. Huamán, *Feature detection and description — opencv tutorials*, https://docs.opencv.org/4.x/db/d27/tutorial_py_table_of_contents_feature2d.html, Accessed: Oct. 21, 2025, 2025.

[37] T. Malche, *What is feature matching? — roboflow blog*, https://blog.roboflow.com/what-is-feature-matching/, Accessed: Oct. 21, 2025, 2024.

[38] S. van der Walt et al., *Skeletonize — scikit-image example gallery*, https://scikit-image.org/docs/0.25.x/auto_examples/edges/plot_skeleton.html, Accessed: Oct. 21, 2025, 2025.

[39] E. I. Education, *What is skeletonization in image processing?* https://www.educative.io/answers/what-is-skeletonization-in-image-processing, Accessed: Oct. 21, 2025, 2025.

[40] B. Tuychiev, *Implementing the dijkstra algorithm in python: A step-by-step tutorial*, https://www.datacamp.com/tutorial/dijkstra-algorithm-in-python, Accessed: Oct. 21, 2025, 2024.

[41] Jenn, *Arc length parameterization (w/ step-by-step examples!)* https://calcworkshop.com/vector-functions/arc-length-parameterization/, Accessed: Oct. 21, 2025, 2022.

[42] Ultralytics, *Ai & computer vision glossary — u-net*, https://www.ultralytics.com/glossary/u-net, Accessed: Oct. 21, 2025, 2025.

[43] A. I. Aramendia. ""the u-net : A complete guide"." Accessed: 30-Oct-2025. [Online]. Available: https://medium.com/@alejandro.itoaramendia/decoding-the-u-net-a-complete-guide-810b1c6d56d8.

[44] Ultralytics, *Ai & computer vision glossary — transfer learning*, https://www.ultralytics.com/glossary/transfer-learning, Accessed: Oct. 21, 2025, 2025.

[45] D. V. Godoy, *Understanding binary cross-entropy / log loss: A visual explanation*, https://towardsdatascience.com/understanding-binary-cross-entropy-log-loss-a-visual-explanation-a3ac6025181a/, Accessed: Oct. 21, 2025, 2022.

[46] R. Vinood, *Dealing with class-imbalanced image datasets*, https://medium.com/data-science/dealing-with-class-imbalanced-image-datasets-1cbd17de76b5, Accessed: Oct. 21, 2025, 2020.

[47] P. Sethuraman, "Optimum assignment and the hungarian algorithm," *Towards Data Science (Medium)*, 2024, Online; accessed 2025-Oct-21. [Online]. Available: https://medium.com/data-science/optimum-assignment-and-the-hungarian-algorithm-8b1027628028.

APPENDIX A

Software Architecture Diagram

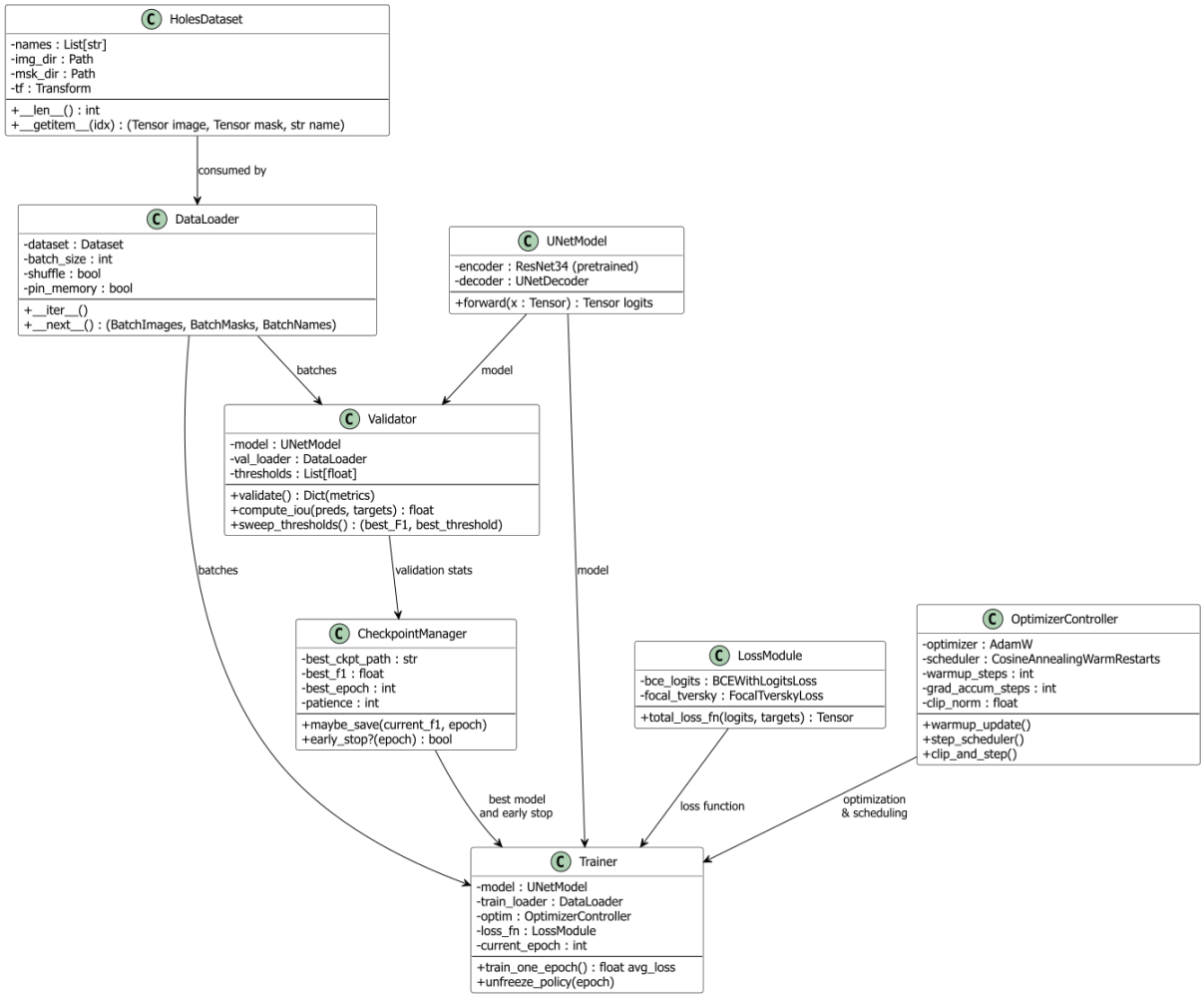
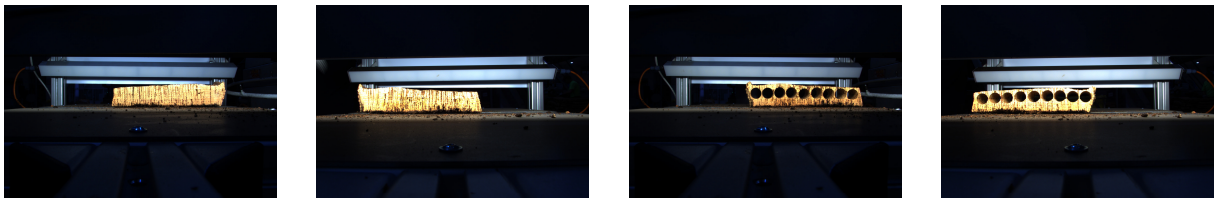


FIGURE A.1. Class diagram representing the main modules of the developed system and their interactions.

APPENDIX B

End-to-End Application Example on a Cork Strip

B.1. Original Images of the Cork Strip



(A) 20240318084533_L (B) 20240318084533_R (C) 20240318084549_L (D) 20240318084549_R

FIGURE B.1. Original images representing the undrilled and drilled views of the strip

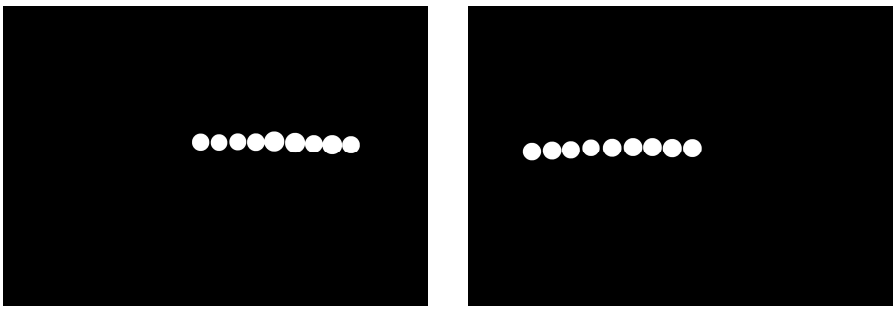
B.2. Cork Strip Detection and Segmentation



(A) 20240318084533_L (B) 20240318084533_R (C) 20240318084549_L (D) 20240318084549_R

FIGURE B.2. Segmented Images of the Cork Strip

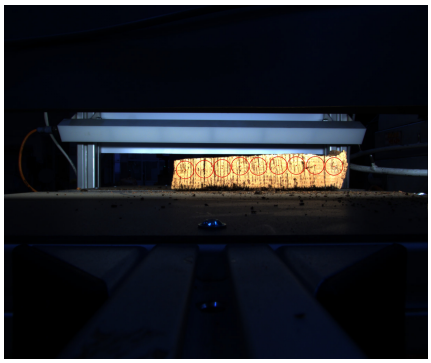
B.3. Holes Detection and Segmentation



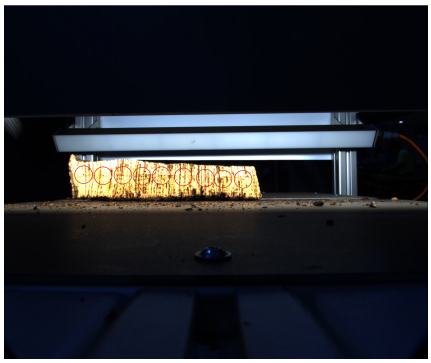
(A) 20240318084549_L (B) 20240318084549_R

FIGURE B.3. Holes Segmentation Binary Mask (drilled images)

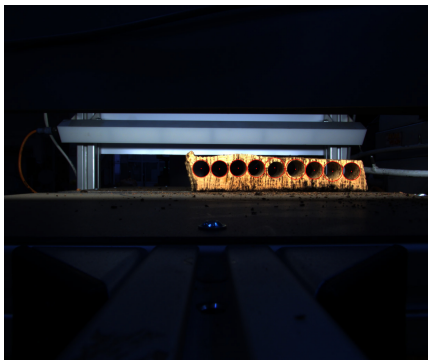
B.4. Holes Projection onto the Undrilled Images



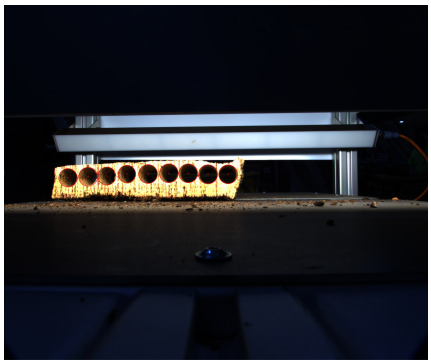
(A) 20240318084533_L



(B) 20240318084533_R



(C) 20240318084549_L



(D) 20240318084549_R

FIGURE B.4. Holes projection onto the undrilled images. Projected hole circumferences marked in red

B.5. "Good Cork" Regions Prediction



(A) 20240318084533_L



(B) 20240318084533_R

FIGURE B.5. "Good Cork" regions predicted in both undrilled images. Contours of the “Good Cork” regions marked in red and circumferences of the ground truth holes marked in green

B.6. Undrilled Images Alignment

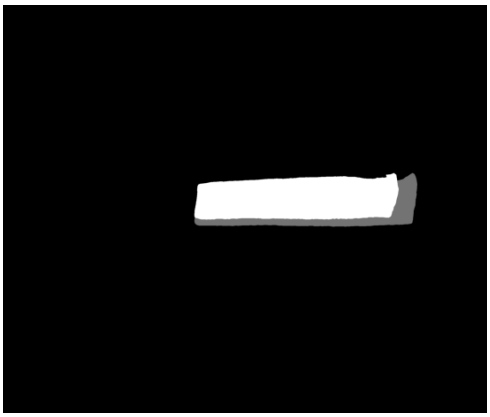


FIGURE B.6. Overlay of undrilled images binary masks after alignment

B.7. Final Predicted Holes

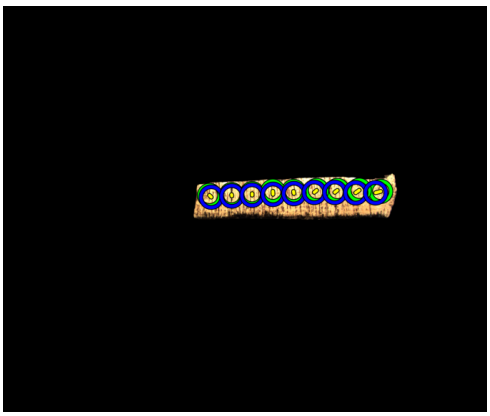


FIGURE B.7. Overlay of predicted holes (after filtering) and ground truth holes. Ground truth holes are marked in green, true positive predictions in blue and yellow lines indicate the displacement between each matched predicted and ground truth hole