DEEP LEARNING-BASED OBJECT DETECTION FOR EFFLORESCENCE RECOGNITION IN MASONRY

October 21st , 2025

MASTER OF SCIENCE ARCHITECTURE, URBANISM AND BUILDING SCIENCES
GRADUATION STUDIO BUILDING TECHNOLOGY

TU Delft

Valentijn Camiel Cloo
5436958

Blank page

Colophon

TITLE	*Exploring the potential of deep learning based image analysis for efflorescence recognition in masonry.*
DOCUMENT	MSc. Graduation Thesis
VERSION	1.0
STATUS	Final
DATE	October 21st 2025

AUTEUR	V.C. (Valentijn) Cloo
STUDENT NR.	5436958
	Delft University of Technology
INSTITUTE	Master of Science Architecture, Urbanism and Building
DEGREE	Sciences
	Architectural Engineering + Technology
DEPARTMENT	Building Technology
TRACK	Julianalaan 134, 2628 BL Delft
ADRES	
	Dr. Barbara Lubelli
1st MENTOR	Associate Professor
TITLE	Heritage & Architecture
CHAIR	
	Dr. Azarakhsh Rafiee
2nd MENTOR	Assistant Professor
TITLE	Geographic Information System Technology
CHAIR	

Preface

This thesis marks the final step in completing my Master of Science in Building Technology at Delft University of Technology. Over the past months, I have had the opportunity to dive deeply into the intersection of heritage preservation and machine learning, a combination that reflects both my passion for cultural architecture and my interest in digital innovation.

I would like to express my sincere gratitude to my mentors, Barbara Lubelli and Azarakhsh Rafiee, for their continuous guidance, valuable feedback, and encouragement throughout the process. Their expertise has played a pivotal role in shaping this research.

As part of this research, I visited several heritage buildings, including the Jesse Church in Delft. Seeing efflorescence and other moisture-related damage up close helped me better understand the practical challenges of masonry preservation. These site visits supported the development and relevance of the method used in this thesis. I'd like to thank everyone who has supported me during these visits.

I also wish to thank my colleagues, friends, and family for their support, patience, and motivation, especially during the more challenging phases of this work.

This thesis is the result of both academic inquiry and personal growth, and I hope it contributes meaningfully to the evolving field of heritage conservation through digital tools.

Valentijn Camiel Cloo
Delft, October 2025

Blank page

Summary

This research explores how deep learning models can improve the detection of efflorescence in masonry buildings in the Netherlands. Efflorescence, caused by moisture-driven salt transport, poses detection challenges due to its variable appearance, similarity to other surface features, and co-occurrence with other forms of masonry damage.

To address this, the study benchmarked two state-of-the-art object detection models: Mask R-CNN and YOLOv8. On a single-class dataset, Mask R-CNN achieved a peak mAP@0.5 of 0.35, outperforming YOLOv8 (0.30–0.33) in segmentation quality and spatial precision. However, both models suffered from false positives, often misclassifying encrustation, lichens, and graffiti as efflorescence due to visual similarity.

To mitigate this, a multi-class training setup was introduced. Graffiti achieved the highest mAP (0.60) and near-perfect precision due to strong visual contrast, while lichens were classified with high stability. In contrast, efflorescence and encrustation remained difficult to separate, resulting in unstable mAP and precision fluctuations over time. This confirmed that misclassification significantly limits model accuracy when damage types share visual characteristics.

Model performance was further evaluated by incorporating thermal imaging (RGBT), combining aligned RGB and infrared data to detect moisture-driven efflorescence. The RGBT model reduced false positives (as few as 3 per evaluation set) and improved detection confidence, reaching a precision of 0.94 and an average confidence score of 0.96, although it required more epochs to converge and showed increased false negatives in ambiguous scenes. Still, RGBT improved the confidence of the visual detection in real-world, poorly lit, or heritage conditions, where over-segmentation is costly.

A spatial co-occurrence analysis of annotated masks indicated a statistically significant correlation between efflorescence and adjacent damage, supporting the potential of dual-class detection

In conclusion, while deep learning models can support efflorescence detection, especially when enhanced with thermal input and multi-class refinement, their performance depends heavily on dataset quality, annotation strategy, and class separability. These findings offer a foundation for scalable, automated inspection in conservation and diagnostics of masonry.

Table of Contents

Abbreviations

MDCS	Monument Diagnosis and Conservation System
MDDS	Masonry Damage Diagnostic System
NDT	Non Destructive Testing
SHM	Structural Health Monitoring
MoSCoW	Must Have, Could Have, Should Have, Won't Have
CNN	Convolutional Neural Network
RCNN	Region-based Convolutional Neural Network
Fast R-CNN	Fast Region-based Convolutional Neural Network
Faster R-CNN	Faster Region-based Convolutional Neural Network
Mask R-CNN	Mask Region-based Convolutional Neural Network
YOLO	You Only Look Once
mAP	Mean Average Precision
FPS	Frames Per Second
RoI	Region of Interest
FPN	Feature Pyramid Network
IoU	Intersection over Union
GPU	Graphics Processing Unit
RGB	Red Green Blue (color image)
IR	Infrared
RGB-T	Red Green Blue + Thermal (image fusion)

Glossary

Term	Definition
Efflorescence	Crystalline deposit of salts on masonry surfaces
Masonry	Construction using individual units (like bricks or stones) bound together with mortar.
Deep Learning	A subset of machine learning that uses multi-layered neural networks to model complex patterns.
Convolution	A mathematical operation used in CNNs to extract spatial features from an image.
Kernel (or Filter)	A small matrix used in convolutions to detect specific patterns (e.g., edges, textures).
Feature Map	The result of applying a convolutional filter to an image, representing learned features.
Pooling	A downsampling operation that reduces the size of feature maps while retaining key information.
Backbone	The main CNN architecture (e.g., ResNet, VGG) used to extract features in an object detection model.
Semantic Segmentation	Assigns a class label to every pixel in an image (e.g., all efflorescence pixels).
Instance Segmentation	Differentiates between individual objects of the same class (e.g., two efflorescence patches).
Bounding Box	A rectangular outline around detected objects in object detection models.
Mask	A binary (or colored) overlay showing the shape and area of detected objects.
Annotation	The process of labeling data (e.g., bounding boxes or masks) used to train and validate models.
Dataset	A structured collection of annotated images used for training, validating, and testing models.
Model Training	The process of teaching a model to recognize patterns by minimizing prediction error over many iterations.
Inference	The use of a trained model to make predictions on new, unseen data.
Thermal Imaging	Technique using infrared sensors to detect temperature differences, helpful in moisture detection.

List of Figures

List of Tables

1. Introduction

Masonry is one of the most commonly used materials in Dutch architecture, especially in older and historic buildings. Its porous nature makes it particularly vulnerable to environmental influences such as moisture ingress and salt crystallization. Efflorescence is one of the most prevalent and visible forms of deterioration which alters visual appearance of facades, it often signals underlying moisture and salt transport processes that can lead to long-term deterioration if not addressed. It occurs when soluble salts migrate to the surface through capillary moisture movement and crystallize upon evaporation, leading to aesthetic damage.

Masonry in the Netherlands is particularly susceptible to moisture-related damage due to its geographic and climatic conditions. Large parts of the country lie below sea level, with high groundwater tables and widespread salinization of water near coastal regions (Deltares & TNO, 2024). These conditions increase the potential for salt intrusion into masonry structures. Additionally, the Dutch climate is increasingly influenced by climate change, with the Royal Netherlands Meteorological Institute (KNMI) projecting drier summers and wetter winters in the coming decades (KNMI, 2023). This fluctuation in moisture levels intensifies the wet-dry cycles that promote salt crystallization and surface decay in masonry (Lubelli et al., 2006; van Hees et al., 2004).

Changes in rainfall patterns, rising temperatures, and urban densification have further contributed to increased instances of dampness and salt-related degradation in facades (Vandemeulebroucke et al., 2023). Next to the effects of climate change, the Dutch construction and heritage sectors also face a shortage of skilled labour. This shortage can lead to improper repairs, the use of incompatible materials, or inadequate diagnosis of moisture problems, further increasing the risk of recurrent damage and reducing the long-term resilience of heritage masonry (Pintossi et al., 2023; Harun, 2011). These factors increase the risk of recurrent damage and reduce the long-term resilience of heritage masonry.

Many buildings in the Netherlands, including churches, monasteries, canal houses, and farms, are decades or even centuries old. These structures often face gradual degradation due to weathering and moisture accumulation. Maintenance budgets are limited, and restoration work is frequently performed reactively rather than preventively. According to a 2021 report by the Dutch Cultural Heritage Agency, only 45.1% of listed buildings were classified as being in 'good' maintenance condition, while 39.3% were rated as 'fair' and 12.1% as 'moderate', underscoring the limited capacity for consistent and optimal preservation efforts *(Erfgoedmonitor, 2021).*

These combined environmental, technical, and societal factors highlight the need for efficient, non-invasive methods to detect and monitor damage, and in particular efflorescence in masonry.

In this chapter, the setup for the research context will be further elaborated. Subsequently, the research objective, research questions, and theoretical framework will be discussed. At the end of this chapter, a reading guide will outline the overall structure of the thesis.

1.1. Research Context

The preservation of historic architectural structures is crucial for safeguarding cultural heritage, Keshmiry & Hassani describe heritage as a vibrant memory of a country's history and development and should be considered to the maximal extent possible (Keshmiry et al., 2024).

Additionally, there have been multiple studies conducted that showcase the role of heritage buildings and their economic value for tourism and local economies (Lazrak et al., 2014), (Koster & Rouwendal, 2017).

A study on the economical value of heritage shows that the daily spending is 60% higher for tourists that where dedicated for cultural heritage. It was also concluded that historic rehabilitation creates 13% higher return on Investment then newly constructed architecture, it holds 16.5% more jobs and produces in order of magnitude 1.2 times less waste (Nypan, 2006). Moreover, built heritage creates positive spillover effects, enhancing the value of real estate in its surrounding areas (Lazrak et al., 2014). It must be noted that the cultural built heritage is valued differently over time. Nijkamp highlights that, following the Second World War, many buildings in the Netherlands were demolished to make way for new developments. However, these historic structures are now highly valued for their social significance (Lazrak et al., 2014).

CHALLENGES OF PRESERVING HERITAGE BUILDINGS

Preserving heritage buildings poses numerous challenges, including weathering due to environmental influences, and the complexity of preservation guidelines. As nearly all conservation projects encompass both repair and maintenance phases, it is crucial for all stakeholders to thoroughly understand building defect diagnostics and apply material treatments with consistent care (Harun, 2011). Additionally, the shortage of skilled workers and experts further complicates preservation efforts, placing the responsibility on conservators to ensure that building practices and materials align with the integrity of the heritage structure (Pintossi et al., 2023).

Over the last decade numerous studies have been conducted on the environmental factors influencing the conservation potential of heritage structures (Hall et al., 2016; Sesana et al., 2018). Sesana et al, conducted an extensive literature review on state of the art impacts of climate change on the built heritage in a broader sense. It was concluded that water is one of the most important decay factors. Rising precipitation levels associated with climate change could lead to soil saturation and the overloading of drainage and runoff systems, thereby increasing the risk of damp infiltration in historic materials, including masonry walls (Sabbioni et al., 2008). Water can also penetrate porous materials through condensation and capillary action, particularly in buildings located in areas where the groundwater level is high. While these processes are not caused by climate change directly, they are exacerbated by it due to fluctuating groundwater levels and longer wet periods. This water ingress accelerates material deterioration by promoting corrosion, biological growth, and salt crystallization within the material (Sabbioni et al., 2007).

Additionally, the high cost and technical complexity of restoring heritage buildings often result in challenges during execution. While preservation guidelines are in place to ensure compatibility and integrity, difficulties arise when these are not properly understood or implemented. This can lead to cost overruns, inappropriate restorations, and extended project timelines, as seen in various large-scale heritage projects worldwide (Roy & Kalidindi, 2017). Davies et al. (2024) highlight the potential of using sustainable natural materials to lower restoration costs and improve effectiveness. However, if not carefully matched with existing materials, such solutions risk incompatibility, potentially undermining long-term preservation goals

MASONRY IN HERITAGE

As a cornerstone of traditional construction, masonry plays a crucial role in Dutch heritage buildings, yet it remains particularly vulnerable to environmental and structural stresses (Cultural Heritage Agency of the Netherlands, 2023). As a material deteriorates over time, its capacity to withstand decreases. Progressive internal cracking and the buildup of damage are key contributors to failure in materials. This behaviour is observed in masonry materials such as bricks, ceramics, concrete, mortar, and stone (Keshmiry et al., 2024).

ADVANCED DIAGNOSTIC TECHNIQUES

Conventional diagnostic methods for heritage buildings, including visual inspection and manual probing, provide only surface-level information and are often unable to detect early-stage or subsurface damage, underscoring the need for advanced diagnostic technologies. In recent years, significant progress has been achieved in analysis, inspection, testing, and monitoring methods for diagnosing masonry heritage structures (Proietti et al., 2021). These advancements are primarily driven by the need for accurate evaluations of structural conditions to ensure the preservation of these structures. Nonetheless, the diagnosis of damage processes affecting heritage buildings remains a complex task (Makoond & Pela, 2021). Besides the studies towards the diagnostics of the current state of the structure there have also been studies that focus on future state of heritage objects in the form of damage functions in relationship with weathering processes. Damage functions are utilized in predictive degradation modelling, but their predictions involve uncertainty because they rely on extrapolation (Strlič et al., 2013).

TECHNOLOGY IN PRESERVATION

Technological advancements have significantly improved the preservation of heritage buildings by enabling non-invasive, accurate, and efficient data collection. Tools such as digital photogrammetry, laser scanning, and drone-based imaging enable precise documentation of geometry, surface decay, and deformation over time (Jung & Mazzetto, 2024). In parallel, the availability of high-resolution cameras and mobile imaging devices has made data collection more accessible and scalable. As a result, large image datasets are routinely generated across building projects. However, much of this data remains underutilized (But, 2024).

Machine learning models, particularly object detection and segmentation algorithms, show strong potential for automating the recognition of damage patterns within large image datasets, improving efficiency and ensuring more consistent assessments across sites (Yu et al., 2025). Recent developments in sensing and machine learning are occurring at a rapid pace, with significant advancements made in just a few years. As a result, preservation technologies must continually adapt to new capabilities. While these tools enhance the speed and efficiency of data processing, challenges remain in maintaining accuracy across varying imaging conditions and damage types. Balancing efficiency with reliability is therefore essential for practical implementation. The developments enable a shift from reactive to preventive maintenance strategies (Mansuri & Patel, 2022). Instead of relying solely on periodic manual inspections, ongoing visual monitoring using automated systems can flag early signs of deterioration, contributing to better-informed, data-driven preservation efforts.

These technologies support preservation professionals in monitoring damage progression and in mapping vulnerable areas, which in turn inform the diagnostic process. Despite this progress, combining different sensing techniques with machine learning remains relatively underexplored in the context of heritage damage assessment, particularly for forms of decay like salt crystallization.

SUSTAINABLE HERITAGE

Preserving existing heritage buildings supports long-term sustainability goals by extending the lifespan of structures, reducing the need for demolition, and minimizing the use of new construction materials. Maintenance and adaptive reuse of heritage sites reduce resource consumption and embodied carbon emissions compared to new construction (Labadi et al., 2021). Despite these benefits, integrating sustainability into heritage management remains inconsistent. Many heritage sites face conflicting regulations, unclear sustainability guidelines, or lack the technical and financial capacity to implement energy-efficient upgrades (UNESCO, 2015). Heritage preservation was long absent from global sustainability agendas, despite its relevance to environmental, social, and economic goals. The 2030 Agenda marked a turning point by formally recognizing culture and heritage as enablers of sustainable development.

RESEARCH GAP

This study addresses the gap between traditional preservation practices based on visual monitoring and the potential of modern technologies to support more accurate and scalable diagnostic processes. Despite growing interest in applying machine learning to heritage diagnostics, current strategies largely rely on manual inspections that are time-consuming, subjective, and difficult to scale.

Although object detection and segmentation models have shown potential in automating damage recognition, their performance often declines under real-world conditions. Factors such as inconsistent lighting, low-resolution images, and viewing angles compromise model reliability and generalizability. These challenges become more critical when attempting to distinguish between damage types with similar visual characteristics, which can lead to misclassification and reduce the practical value of automated methods.

While sensing tools such as photogrammetry and thermal imaging are becoming more accessible, they are rarely integrated into machine learning workflows. Research remains limited on how these techniques can complement RGB-based analysis, especially when image quality is suboptimal. This is particularly important for decay types like efflorescence, which may resemble other surface-level discoloration.

In addition, the ability of machine learning to support spatial interpretation, such as detecting patterns in the proximity or co-occurrence of efflorescence and damage, has received little attention. Investigating these spatial relationships can provide deeper insight into deterioration processes, yet remains an underexplored aspect of heritage diagnostics.

1.2. Problem Statement

Recent years have seen significant advancements in analysis, inspection, testing, and monitoring techniques for diagnosing damage in masonry heritage structures. These developments are driven by the need for accurate evaluation of the current condition to ensure the preservation of these monuments. Despite these innovations, diagnosing damage in heritage buildings remains a challenging task, particularly when dealing with unique and complex structures. The difficulties arise from the interaction between different structural components (e.g., walls, foundations, and facades) and the mechanical, physical, and chemical properties of the materials.

Visual assessment plays a critical role in identifying damage in masonry. Among the various types of damage, efflorescence, a visually detectable form of salt deposition, represents a significant challenge for reliable diagnosis due to its impact on both appearance and material durability. Moisture-related damage, including efflorescence and discoloration, compromises the durability of masonry materials. Traditional methods for detecting damage rely on manual visual inspection, which is time-consuming, labor-intensive, and prone to subjectivity.

Advancements in deep learning, particularly Convolutional Neural Networks (CNNs), have enabled promising developments in image-based damage detection. Techniques such as semantic segmentation and classification are commonly employed to identify specific damage patterns. However, accurately detecting and quantifying visually evident damage like efflorescence using automated methods remains a complex task due to variations in surface textures, lighting conditions, and the subtle nature of the damage itself. This research seeks to evaluate the performance of a known machine learning model in detecting efflorescence and explore enhancements to improve its accuracy and reliability, contributing to more efficient and effective damage assessment methods.

1.3. Research Objective

This study aims to bridge the gap between traditional survey techniques and modern technology by leveraging advanced machine learning methods for the accurate detection of efflorescence in masonry heritage buildings Specifically, it seeks to evaluate the capabilities of a pre-existing deep learning framework, based on Convolutional Neural Networks (CNNs), to detect efflorescence under varying real-world conditions. While the focus is on efflorescence, other damage types such as discoloration, encrustation, biological growth, and graffiti are considered insofar as they cause misclassification risks or co-occur with efflorescence. Furthermore, the research aims to identify and implement enhancements to improve the model's accuracy and reliability, addressing challenges posed by variations in surface textures, lighting conditions, and the distinction from look-alike damage types, such as encrustation, some forms of biological growth, and graffiti. A key objective is to assess whether incorporating thermal imagery can improve detection accuracy under varying real-world conditions. Through this approach, the study contributes to the development of efficient, automated methods for visual damage assessment in heritage preservation.

1.4. Research Questions

To accomplish the intended goal, the following central research question was formulated:

How can deep learning models be applied to improve the detection of efflorescence in masonry buildings in the Netherlands?

The following sub-questions are designed to support answering the main research question:

SQ1: What are the visual characteristics of efflorescence on masonry, and how do these factors present challenges for detection?

SQ2: Which deep learning models are most suitable for detecting and classifying efflorescence on masonry, based on performance criteria?

SQ3: What is the effect of variables (such as image quality, lighting, and orientation) on the performance of the model?

SQ4: How can the performance be improved by addressing misclassification of similar damage types and co-occurrence with efflorescence?

SQ5: How can the integration of thermal (IR) imagery improve the detection accuracy and reliability of efflorescence in masonry?

SQ6: How well does the enhanced model perform when evaluated on unseen data and applied to real-world case studies of efflorescence?

1.5. Research Design

This study is structured into four sequential phases, as depicted in: figure 1, (1) Literature Research, (2) Experimental Design, (3) Validation & Application and (4) Reporting of the analysis.

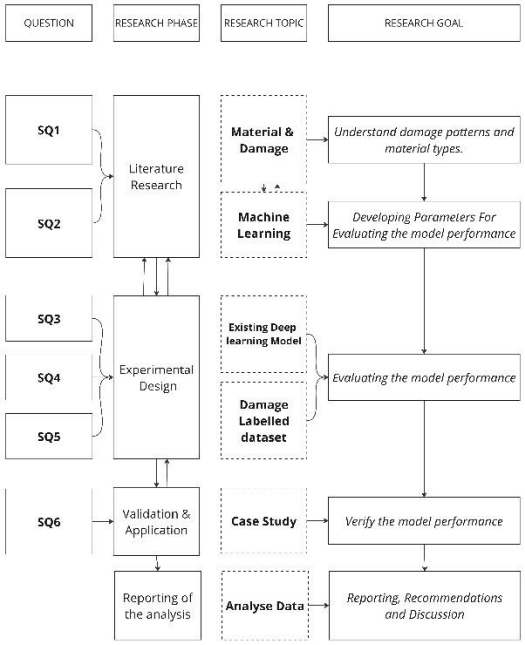


Figure 1 Research Design Overview

Literature Research

The Literature Research phase consists of two main topics that will be researched The first component establishes a broad understanding of masonry materials and damage processes, damage mechanisms related to efflorescence. The objective of literature research is to get a better understanding of damage patterns and material types. This research forms a basis for developing the general outline of the experiments and what metrics could be used to evaluate the models performance. The characteristics and history of masonry will be discussed following the damage processes and causes to the masonry structures. For the classification of the damage types, references will be made to the MDDS[1] or MDCS[2] hosted by TNO, TU Delft and the Cultural Heritage Agency of the Netherlands due to their significance in the field of heritage damage diagnostics.

The second component introduces the theoretical foundations of state-of-the-art computer vision models and their application in the field of heritage based computer vision aided diagnostics. Existing machine learning models are analysed and classified and their differences are discussed. Core

[1] Masonry Damage Diagnostics System
[2] Monument Diagnosis and Conservation System

concepts such as convolutions, feature maps, and activation functions are explained at an accessible level to clarify model behaviour without advanced mathematics. The required datasets and labels are reviewed and the limitations of the models identified to point out knowledge gaps to be addressed in the research.

Experimental Design

The experimental design builds on insights from literature research, expert consultation, and preliminary model considerations. The first step will establish a performance benchmark by evaluating two pre-selected models, YOLO and Mask R-CNN, on a representative dataset. This benchmark is intended to identify baseline performance and potential limitations. The design of the experiments is informed by both theoretical findings and practical constraints, ensuring that they are feasible, relevant, and aligned with the available time and scope. The experiments are structured in multiple steps, each targeting a specific research hypothesis and progressively building toward improved model performance.

Several working hypotheses are formulated from literature and anticipated challenges: (1) the impact of integrating thermal imaging on detection accuracy of efflorescence; (2) the influence of variable image conditions (lighting, angle, scale) on robustness; (3) the extent of misclassification between visually similar damage types, such as graffiti and biological growth; and (4) the spatial co-occurrence of damage and efflorescence in masonry. Each hypothesis guides a dedicated set of experiments designed to answer targeted sub-questions.

The first experimental phase will involve a benchmark evaluation of two pre-selected models: YOLO and Mask R-CNN. These models are selected based on their tested repositories and relevance in literature. Performance will be assessed using metrics such as accuracy, precision, recall, and F1-score. Based on these results, the model that demonstrates higher spatial precision will be advanced for further experimentation. This benchmark established the model's initial capabilities and helped identify areas for improvement.

Subsequent experiments will incorporate additional parameters to simulate real-world variability and enhance robustness. These include image quality factors (e.g., resolution, contrast), environmental conditions (e.g., oblique angles, uneven lighting), and the integration of thermal imagery to support moisture detection. Unlike studies that rely on controlled datasets with orthogonal, fixed-distance images, this research will employ a diverse dataset with high variation in capture conditions. Each parameter will be tested individually and in combination. Based on outcomes, the model will be fine-tuned through transfer learning, hyperparameter optimization, and data augmentation, with the aim of increasing accuracy while limiting false positives.

Evaluation will include segmentation accuracy for detecting damage at the pixel level, classification accuracy for distinguishing damage types (e.g., efflorescence vs. biological growth), processing time for real-time feasibility, and robustness across varied datasets. Additional analysis will examine misclassification among visually similar categories and assess the spatial relationship between damaged and efflorescent bricks.

The dataset used in this study comprises heritage damage images with substantial variation in quality and perspective. To ensure relevance, images were manually reviewed and labeled at the brick level to improve semantic consistency, with annotations validated where necessary through consultation with domain experts.

The experimental phase follows an iterative approach, where interim results will guide refinements in later experiments. For example, thermal imaging will be integrated if RGB-only detection proves insufficient under field conditions, and hypotheses concerning spatial co-occurrence or misclassification may be added as dataset challenges emerge. This flexible, adaptive structure ensures continuous improvement throughout the study.

Validation & Application

The Validation and Application phase assesses the performance of the enhanced machine learning model using a real-world case study. A representative masonry wall (approximately 15–20 meters in length) with known instances of moisture-related damage was selected. This wall featured similar materials and damage characteristics as those found in heritage structures. A point cloud scan and high-resolution image dataset were captured for this wall, serving as the foundation for applying the trained model. While thermal imagery was also collected at the site, technical limitations prevented full alignment of thermal data across the entire wall surface. As such, thermal analysis was limited to image-level inference rather than full-scene application. Validation was carried out using standard unseen validation sets held out during training, ensuring a consistent and objective evaluation of model performance.

The model was applied to this real-world dataset to reproduce findings from the experimental phase and evaluate its robustness under practical conditions. Images captured from the selected wall were analyzed using the trained model to detect and classify instances of efflorescence and related damage. The outcomes were compared to manual visual observations made on-site to assess the accuracy of the predictions. Although external expert validation was not conducted, internal model validation was ensured through separate, held-out datasets during training, reflecting performance on unseen data.

A representative masonry wall was selected from a heritage site located at the Vijzelgracht in Amsterdam, part of a 19th-century cloister complex originally built as a women-only monastery. The site features an inner courtyard with exposed brick façades that display visible moisture-related damage such as efflorescence and discoloration. This location was chosen for its historical relevance, material similarity to other heritage masonry structures, and accessibility for data collection.

Reporting & Analysis

The Reporting & Analysis phase serves as the finalization of the research, consolidating all findings, insights, and evaluations from the previous phases into a complete document. This phase ensures that the research outcomes are presented clearly and meaningfully, while also providing recommendations for future research and practical applications.

1.6. Reading Guide

The structure of this thesis is as follows: Chapter 2 provides a literature review focused on the general methodology for masonry damage diagnostics and the potential application of machine learning in this field. Chapter 3 introduces the methodology for the experiments. Chapter 4 outlines the hypotheses for each experiment and summarizes the results. Chapters 5 and 6 present the discussion and conclusion, along with recommendations for further research. Finally, Chapter 7 reflects on the overall research process, examining what could have been done differently and identifying the key factors that contributed to or hindered the success of the thesis.

2. Literature Research

The first step in this review involved a systematic literature search using Scopus according to the PRISMA systematic review (Fino et al., 2023) and the outline is further described in **figure 2** *Literature research process*. The databases were queried using Boolean operators to combine groups of keywords[3] related to:

1. Technique ("machine learning" OR "deep learning"),
2. Target ("heritage buildings" OR "historic buildings"), and
3. Purpose ("damage detection" OR "diagnosis" OR "assessment").

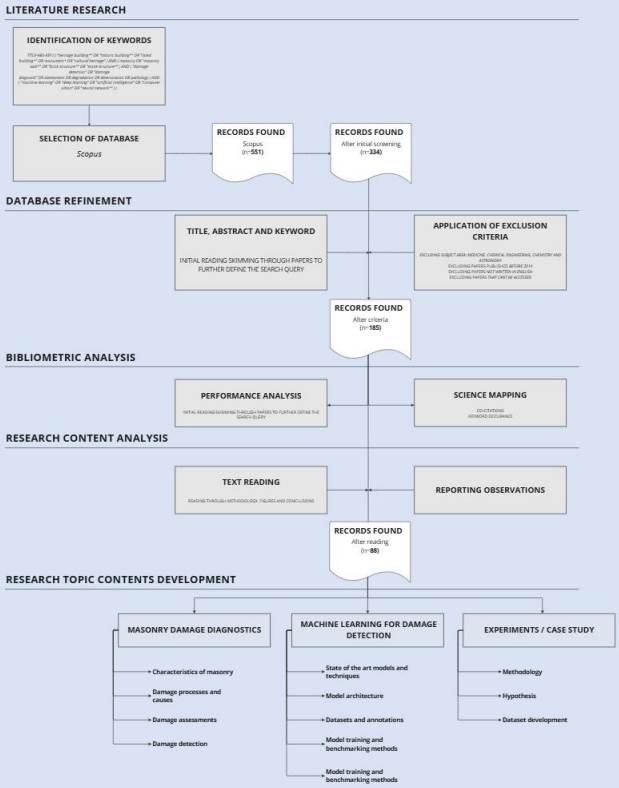


Figure 2 Literature research process adapted from (Fino et al., 2023).

This approach aimed to capture studies focusing on the use of advanced digital or computational methods (technique) applied to historic or heritage structures (target) for the purpose of detecting, diagnosing, or assessing deterioration (purpose). The search used a topic-dependent time window. For machine-learning/object detection, results were limited to 2014–2024 to capture recent developments. For efflorescence and masonry-diagnostics literature, no strict lower bound was applied; seminal works prior to 2010 were explicitly included via targeted queries and backward snowballing. The search was restricted to English and Dutch-language publications (research articles, review papers, conference proceedings, and book chapters). After merging and de-duplicating results, 551 records remained for screening.

[3] *The final search query defined as: TITLE-ABS-KEY (("heritage building*" OR "historic building*" OR "listed building*" OR monument* OR "cultural heritage") AND (masonry OR "masonry wall*" OR "brick structure*" OR "stone structure*") AND ("damage detection" OR "damage diagnosis" OR assessment OR degradation OR deterioration OR pathology) AND ("machine learning" OR "deep learning" OR "artificial intelligence" OR "computer vision" OR "neural network*"))*

In the second phase, abstracts and author keywords were examined to verify that each document met predefined inclusion criteria and could be accessed with the open access institution license, such as: (1) A clear focus on masonry or heritage building materials (rather than general construction or modern materials), (2) explicit discussion of damage, degradation, or assessment (not purely design or architectural history), (3) use or proposal of a machine-learning or data-driven approach. Any study not meeting these criteria, for instance, articles limited to new building designs or purely theoretical machine learning models without any masonry application were excluded from further analysis. As a result of this filtering, **185** articles remained.

To gain insights into the thematic structure and research trends, the selected references were imported into VOSviewer for bibliometric and keyword co-occurrence analysis. VOSviewer facilitated the visualization of clusters, indicating the key concepts as shown in **figure 3:** *Co-occurance of keywords < 5 of a maximum of 100 words.*

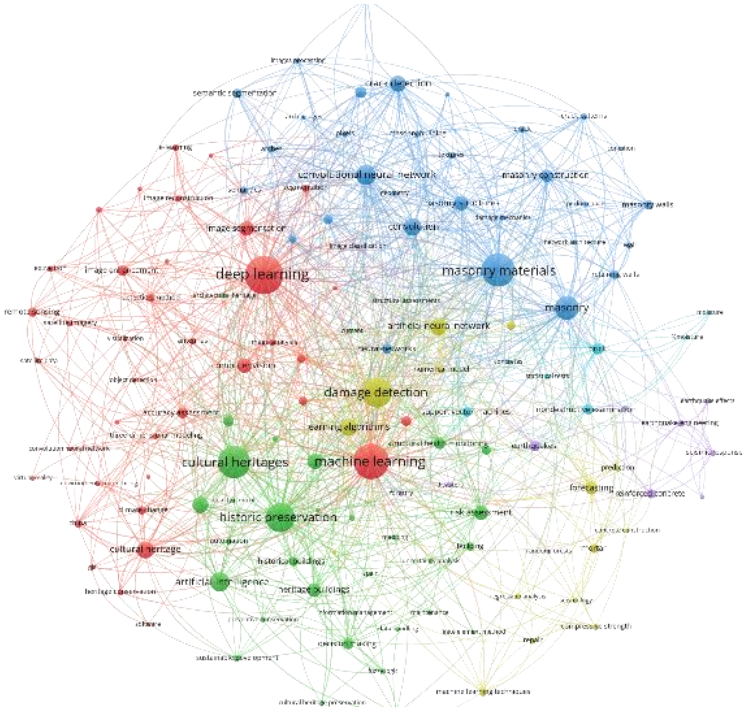


Figure 3 Co-occurrence of keywords < 5, a selection of max 100 words.

The keyword mapping helped further develop the literature research framework and the contents. Regarding damage diagnostics and masonry, terms such as “efflorescence,” “moisture,” “cracks,” and “brick façade” frequently clustered together, indicating the centrality of these issues in current masonry research. Keywords like “deep learning,” “neural networks,” “semantic segmentation,” and “image analysis” often formed a distinct group, showing an emerging focus on automated detection methods.

Few studies explicitly tied advanced imaging (like thermal or NIR) to machine learning frameworks for diagnosing moisture-induced damage in detail. This indicates an opportunity to combine different methods. Although general “masonry damage” appeared often, certain niche issues, like micro-cracks in mortar or salt crystallization at varying environmental conditions, showed fewer connections, hinting at potential gaps worth exploring. Keyword clusters can inform which subtopics merit deeper investigation. For example, a high co-occurrence of “deep learning” and “cracking” suggests that this area is well-studied and might already have mature solutions. On the other hand, “mortar damage” rarely appears alongside “machine learning,” it might indicate a gap for new research.

2.1. Masonry Damage Diagnostics

The conservation of heritage buildings requires a comprehensive approach to damage diagnostics and risk assessment. Several internationally recognized methodologies have been developed to address this need. The Building Condition Audit, as outlined by the British Standards Institution (BS 7913: 2013), provides a systematic framework for evaluating the physical state of historic structures. The Cultural Heritage Risk Assessment Model (CHARM), developed by ICCROM in collaboration with the Canadian Conservation Institute, offers a holistic approach to identifying and mitigating risks to cultural heritage (Michalski et al., 2016). UNESCO's guidelines on heritage conservation emphasize the importance of regular monitoring and assessment to prevent deterioration (UNESCO, 2024). Additionally, the Institute of Historic Building Conservation (IHBC) has provided guidance on retrofitting historic buildings, balancing the need for modernization with the preservation of cultural significance (IHBC, 2021). Within this broader context of heritage conservation methodologies, the specific field of masonry damage diagnostics has emerged as a crucial area of study, given the historical significance of masonry in architectural heritage

The field of masonry damage diagnostics has evolved significantly since the foundational work of (van Hees et al, 1995) and (Van Balen, 1998)), who developed systematic approaches for evaluating deterioration in historic brick structures based on decision tables. Van Hees et al. introduced the Masonry Damage Diagnostic System (MDDS[4]) in a collaborative EU effort, an expert system for assessing ancient masonry, while Van Balen contributed to the creation of a comprehensive damage atlas. Building upon these early frameworks, recent research has focused on integrating advanced technologies and methodologies. These developments include improved non-destructive testing techniques, representing a significant leap forward in the field's capabilities for early detection and prevention of damage. **Figure 4** illustrates the basic concept of Masonry Damage diagnostics.

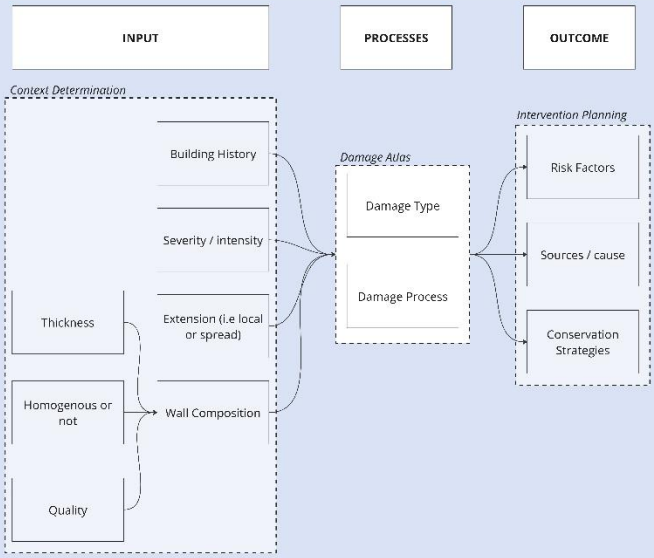


Figure 4 Masonry Damage Diagnostics framework adapted from (R. van Hees et al., 2009; R. P. J. van Hees & Naldini, 2020)

This chapter provides an in-depth understanding of masonry, common damage types, and the existing diagnostic systems used to assess them. It builds a foundation for exploring machine learning applications in damage detection.

[4] The MDCS developed during the COMPASS project is based on the original MDDS but since more materials were added later on, it was renamed to Monument Damage Diagnostic System (R. van Hees et al., 2009)

2.1.1. Characteristics of Masonry in the Netherlands

Brick masonry has long been central to Dutch construction, largely due to the abundance of clay in river delta regions. Bricks were made from locally sourced clay, often mixed with sand or other materials to improve shaping and performance. These materials, while effective, also resulted in porous masonry that is sensitive to weathering and salt-related decay.

Brickmaking was introduced in the Netherlands during the 12th century by monastic communities and became widespread due to material availability and ease of production. Techniques evolved over time, from hand-formed kloostermoppen to standardized industrial bricks, shaping both rural and urban buildings still present today.

Masonry styles varied regionally based on local resources and architectural trends. Northern provinces favored large, robust bricks for religious structures, while river regions used smaller, fine-textured bricks like IJsselsteentjes. In urban areas, decorative and machine-made bricks emerged during industrialization. These differences illustrate the diverse heritage of Dutch masonry and the range of materials involved in preservation today.

2.1.2. Damage Types and Processes

To assess the extent of damage, it is essential to first establish a clear definition. Damage can be broadly understood as an alteration in condition, decrease in worth, or loss in functionality or performance (SA Smith, 2011). The concept of damage relates to various aspects, including physical deterioration, economic losses, and functional impairments (Korswagen et al., 2024). From a technical perspective, damage might focus on structural or material changes.

To continue, (Lourenço et al., 2014) provides a comprehensive definition of degradation and damage in the context of masonry structures: "The degradation processes (chemical, physical and mechanical) exert stresses on the materials, which weaken the material until it fails and damage becomes visible. Degradation can be defined as an increase in decay, which corresponds with a decreasing performance of the material. Thus, damage can be defined as an unacceptable reduction of the performance of the material, affecting its durability."

As previously mentioned in chapter **2.1** the study of masonry damage diagnostics has been significantly advanced by the work of van Hees and van Balen, who developed foundational systematic approaches for evaluating deterioration in historic brick structures (R. van Hees et al., 2009; Van Balen, 1998).

Masonry structures are vulnerable to a wide spectrum of environmental influences and internal deterioration processes. For clarity and consistency, this section distinguishes between environmental factors (external agents), damage processes (decay mechanisms), and damage types (observable manifestations or decay types). The focus of this thesis is on material deterioration driven by environmental factors, particularly moisture-related phenomena. The classification below is based on the work of Korswagen et al. (2024) and Lourenço et al. (2014), supplemented with diagnostic insights from the MDCS framework.

- **Environmental factors**: These are external influences that trigger or accelerate decay processes. Examples include rain, groundwater, air humidity, pollution (e.g. SO_2), and temperature fluctuations (including frost). The presence of moisture—through rising damp, seepage, or condensation, often initiates or sustains damaging processes.
- **Damage processes**: These are the underlying decay mechanisms that reduce the performance of masonry. Moisture-related processes are particularly impactful, and include:

- **Salt crystallization**, where soluble salts (e.g. chlorides, sulfates) crystallize near or within pore walls, generating internal stress and leading to powdering, spalling, or exfoliation.
- **Chemical conversion**, such as the transformation of lime into gypsum or the formation of swelling salts (ettringite, thaumasite), which may lead to blistering, bulging, or bursting.
- **Frost damage**, caused by the expansion of freezing water in saturated materials, leading to delamination or cracking through ice-lens formation.

- **Damage types**: These are the physical symptoms resulting from decay processes. Common types include:
  - **Disintegration**: Powdering, crumbling, scaling.
  - **Layering and exfoliation**: Especially in frost-damaged or salt-contaminated mortar.
  - **Surface changes**: Staining, efflorescence, crust formation, or graffiti.
  - **Mechanical deformation**: Bulging or displacement.

This refined classification not only aligns with the MDCS system but also helps distinguish damage-processes from types, which is essential for diagnosis and model development.

The Monument Diagnosis and Conservation System (MDCS) provides a framework for classifying and analyzing damage processes and types, particularly in heritage structures. Originally based on the MDCS web application developed by TNO, TU Delft, and the Dutch Rijksdienst voor het Cultureel Erfgoed the main masonry related damage types are summarised in **Table 1:**

Damage Type	Sub-Type	Sub-Sub type
Brick Surface Change		
	Chromatic Alteration	
		Moist spots
		Staining
	Deposit	
		Soiling
		Graffiti
		Encrusation
		Efflorescence
		Crypto-florescence
	Transformation	
		Patina
		Crust
Disintegration		
	Detachment	
		Loss of Adhesion
		Blistering Paint
		Peeling Paint
	Loss of Cohesion	
		Powdering
		Crumbling
		Brick-blistering
		Erosion
		Cratering
		Alveolization
	Layering	
		Delamination
		Exfoliation
		Spalling
		Scaling
Cracking		
	Crack	
	Haircrack	
	Crazing	
	Star Crack	
Deformation		
	Bending	
	Bulging	
	Displacement	
Mechanical Damage		
	Scratch	
	Incision	
	Perforation	
	Splitting	
	Chipping	
Biological Growth		
	higher Plants	
	Lichens	
	Liverworts	
	Algea	
	Mosses	
	Moulds	
Missing Material		

Table 1 Damage Atlas classification of Brick damages (retrieved from Bonduel, M. (2020))

2.1.3. Damage Diagnostics

Visual inspection[5] is a fundamental method in diagnosing masonry damage, serving as an first step in evaluating the condition of structures. This non-invasive approach involves the systematic observation and documentation of visible defects on buildings. However, visual survey inspection requires skilled personal and there is a possibility for misinterpretation of the results (Gonçalves et al., 2015). In practice, visual assessments are often supported by both destructive and non-destructive investigation techniques to achieve a more complete understanding of masonry damage. Non-destructive testing (NDT) methods, in particular, can complement visual observations by offering insights that are not immediately visible. Techniques such as infrared thermography (IRT) and photogrammetry, for example, can be integrated into photographic surveys to validate or challenge initial hypotheses about the type or extent of damage. These methods are especially useful for detecting issues like hidden moisture, material heterogeneities, or surface deformation patterns. Infrared thermography IRT is effective for assessing grout placement and can identify areas of moisture infiltration or thermal bridging (Alexakis, Delegou, Mavrepis, …, et al., 2024).

Van Hees emphasizes the significance of long-term monitoring in the assessment of masonry damage, particularly for moisture-related deterioration. His research underscores the value of combining periodic visual inspections with environmental data, such as moisture content and salt presence to support a more complete and reliable diagnosis of material degradation over time (R. P. J. van Hees & Naldini, 2020). Within the scope of this thesis, such methods are relevant not only for surveying the current condition of masonry, but also for tracking the progression of damage. When paired with photographic and non-destructive techniques, they can help substantiate hypotheses about the nature and causes of deterioration in heritage structures.(R. P. J. van Hees & Naldini, 2020).

2.1.4. Challenges of damage detection

Detecting damage in masonry structures involves several challenges::

- Complexity of heritage structures: Heritage masonry buildings often have complex geometries, non-homogeneous materials, and different construction techniques, making imaged based damage assessment difficult (Soleymani & Jahangir, 2023).
- Limited information: The absence of design and construction documentation, as well as lack of data on materials used, complicates the assessment process .
- Non-destructive testing requirements: Due to the historical value of these structures, destructive testing methods are often prohibited, limiting to non-destructive testing (Soleymani & Jahangir, 2023).
- Need for high-precision monitoring: Detecting small changes in structures requires high-detailed displacement sensors and advanced monitoring systems (Korswagen et al., 2024).
- Variability in damage: Certain types of damage, such as efflorescence, may only be visible under specific conditions (e.g., dry periods), making consistent detection challenging (Alexakis, Delegou, Mavrepis, Rifios, et al., 2024).
- Multifaceted nature of deterioration: masonry structures often suffer from multiple types of damage simultaneously, requiring a combination of detection and monitoring techniques (Gonçalves et al., 2015).

[5] *Visual inspection* in this context refers broadly to the observation of visible damage characteristics on surfaces, whether through in-person site assessments or the analysis of imagery (e.g. photographs, scans) that capture surface conditions

2.1.5. Salt-Induced Deterioration: Goal and Scope Definition

This research focuses specifically on the assessment of efflorescence in masonry, a form of salt crystallization damage that is frequently observed in historic buildings. Efflorescence is selected as a target due to several key reasons:

- It is a recurring and widespread form of salt-related damage in masonry.
- It is visually and photographically detectable, making it suitable for automated or semi-automated assessment.
- A large archive of images documenting efflorescence is available, supporting data-driven analysis and training of detection models.

Efflorescence is defined in the MDCS (Material Degradation Classification System) as:

"A visible deposit of white salts on the surface of porous building materials, resulting from the migration of salt-laden moisture to the surface and subsequent evaporation."

Salt-related deterioration in masonry originates from the movement and crystallization of soluble salts within the pore network of materials. When water containing dissolved salts migrates through porous media like brick or mortar, it eventually evaporates, leaving behind solid salt crystals. The location of this crystallization plays a crucial role in determining whether the result is purely aesthetic or damaging.

- If salts crystallize on the surface, they form efflorescence, which is generally not harmful to the structural integrity of the material, though it can be visually disturbing.
- If salts crystallize within the pores of the material, they can exert significant pressure on the pore walls, potentially exceeding the material's tensile strength. This crypto-efflorescence often results in mechanical damage such as spalling, scaling, powdering, or loss of cohesion (Lubelli et al., 2004; Lourenço et al., 2014).

Soluble salts dissociate into positive and negative ions in the presence of water. These ions migrate with the moisture and can later recombine to form crystalline solids as water evaporates (Nijland et al., 2018). All salts have the potential to crystallize either on the surface as efflorescence or within the pores of the material, depending on factors such as moisture supply and drying rate. Some salts are more prone to crystallize at the surface than others, while others more frequently precipitate internally, where they can exert significant crystallization pressure. The specific solubility and crystallization properties of each salt type largely determine whether the outcome is a visible deposit or internal material damage. Nijland et al. (2018) highlight the distinction between salts that cause visual deposits and those that actively damage materials.

The crystallization process is dynamic and influenced by environmental conditions such as humidity, temperature, and the availability of water. Repeated wetting and drying cycles can lead to progressive degradation, particularly in historical masonry, where traditional materials often have lower resistance to crystallization pressure (Lopez-Arce et al., 2009).

In this context, the research aims to better understand and detect efflorescence as a proxy indicator for underlying salt activity. While efflorescence itself is not damaging the material, its presence may indicate salt movement within the wall system and help identify areas at risk of deeper deterioration. Thus, it serves as an accessible marker for broader salt-induced decay processes in masonry.

2.1.5.1. Efflorescence in the Netherlands

Efflorescence is a common and recurring phenomenon in Dutch masonry, particularly in coastal, urban areas and rural areas due to the use of fertilizer and excrements of animals can lead to the presence of nitrates salts. It results from the crystallization of soluble salts on the surface of masonry, often forming visible white deposits. This type of salt-induced decay is especially relevant in the Netherlands due to a combination of construction practices and environmental exposure.

In coastal regions, salts were often introduced during construction, especially when locally sourced dune sand (duinzand), naturally rich in chlorides, was used as an aggregate in mortar. This material choice, combined with the maritime climate and its seasonally high humidity and salty air, contributes to continued salt exposure through wind-driven deposition and moisture absorption. Molenaar (2021) notes that chloride contamination can be found in masonry structures located up to 10 km inland from the coast.

In urban settings, additional chloride sources stem from winter road maintenance. De-icing salts such as sodium chloride (NaCl) and calcium chloride ($CaCl_2$) are commonly used to prevent ice formation on streets and pavements. These salts are often redistributed onto adjacent masonry by traffic splash or wind (Steiger et al., 2011). Over time, they penetrate brick and mortar and accumulate within the porous structure of masonry walls. The repeated freeze-thaw cycles that occur in winter further amplify the damage, as absorbed moisture expands upon freezing, creating internal stress (Charola & Bläuer, 2015a). Prolonged exposure to road salts can also accelerate the corrosion of embedded metal elements such as anchors or cavity ties, leading to structural degradation (Molenaar, 2021).

The occurrence of efflorescence is of particular relevance to this study because it is easily detectable through visual inspection and photographic surveys. This makes it a practical and consistent damage type for automated detection using image-based techniques. As this research involves the development of a custom dataset, understanding the locations and conditions under which efflorescence commonly appears, such as near ground level, close to roads, or in coastal zones, helps guide targeted data collection and annotation.

2.1.5.2. Detection of efflorescence

Efflorescence, most commonly recognized as a white deposit on a dry surface, can be detected with the naked eye. However, assessing its severity and identifying the underlying causes remain challenging. To gain insight into the methodologies used in practice for efflorescence detection and analysis, I reached out to several façade renovation companies in Delft and The Hague. Through these discussions, it became evident that there is no standardized assessment approach incorporated in their workflow. Instead, evaluation methods vary across companies. While some rely solely on visual inspection, others incorporate basic moisture measurements. Similarly, treatment strategies are highly case-dependent, influenced by factors such as the extent of the damage, masonry type, and environmental conditions.

Distinguishing efflorescence from visually similar phenomena is another recurring challenge in practice. Surface deposits such as graffiti, encrustations, and molds can resemble efflorescence in color or distribution. Graffiti, while anthropogenic and unrelated to material decay, often appears as light surface layers that confuse detection. Encrustations, typically composed of calcium carbonate or other mineral deposits, share a similar white tone and texture. Molds, categorized under biological growth in the MDCS, may also appear as whitish surface films, particularly in damp conditions. These

look-alike phenomena as seen in figure 5, are not damage mechanisms in the structural sense, but they increase the risk of misclassification during visual or image-based surveys.

Figure 5 Examples of surface phenomena visually similar to efflorescence on Dutch masonry: (left) biological growth (lichens/mold) on a historic brick façade, (middle) encrustation deposits on a quay wall in Amsterdam, and (right) graffiti on a masonry wall in Delft

Nonetheless, a common trend emerged: when efflorescence persists despite cleaning or surface treatments, contractors often resort to extensive masonry renovation. This typically involves removing and replacing affected bricks and mortar joints, or in some cases, rebuilding entire facade sections. Yet even such comprehensive interventions do not always prevent recurrence, particularly if the underlying moisture and salt transport mechanisms are not adequately understood or resolved.

A widely used standard for assessing salt contamination in masonry is the Austrian ÖNORM B 3355-1 "Trockenlegung von feuchtem Mauerwerk – Bauwerksdiagnostik und Planungsgrundlagen". This standard is often used as a guideline in practice for diagnosing and remediating moisture-related damage in buildings. The standard enjoys broad acceptance in practice, likely due to its stringent assessment criteria, which ensure a thorough evaluation of salt-related damage (Snepvangers, 2005). While it enjoys broad acceptance, the standard also has notable limitations. It provides general threshold values for salt concentrations but does not account for the specific effects of different salt compounds, which may vary considerably in their damaging impact. Furthermore, the methodology does not explicitly address the role of sampling depth, even though salt accumulation is often concentrated in the outer millimetres of masonry. These simplifications can influence the accuracy of assessments and may lead to misinterpretation of the severity or risk of salt-induced damage.

To continue, The WTA-Merkblatt 4-5-99/D "Beurteilung von Mauerwerk" provides a structured approach for assessing masonry, outlining six key steps: (1) Orientation, (2) Recording and classifying damage, (3) Investigation planning, (4) On-site and laboratory investigations, (5) Evaluation of investigation results, and (6) Restoration plan. This systematic methodology serves as a foundation for diagnosing and addressing moisture related masonry damage. In practical applications, similar assessment frameworks are often followed. One such approach is outlined in **figure 6**, which provides detailed flowchart illustrating the sequential steps from building analysis to the implementation of measures.

Moisture measurement in masonry is essential for assessing salt-induced damage, diagnosing the root causes of deterioration, and determining appropriate remediation strategies. Various parameters are used to evaluate moisture behavior in masonry, as described by Snepvangers (2005). These include (1) moisture content, expressed in volume percentage, mass percentage, or using the carbide method (CM-%), (2) degree of saturation, (3) maximum water absorption capacity, (4) critical moisture content, (5) hygroscopic equilibrium moisture content, (6) water absorption coefficient, and (7) water penetration coefficient. These parameters provide valuable insights into the moisture dynamics of masonry and their role in efflorescence formation (Snepvangers, 2005).

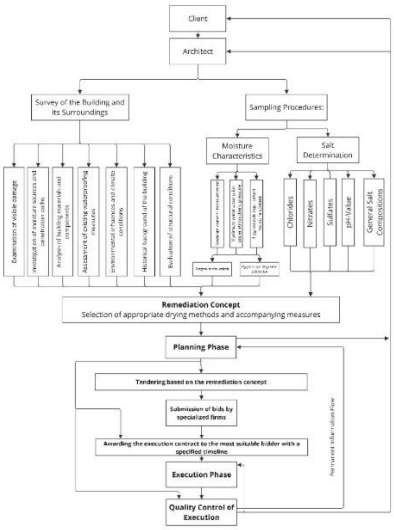


Figure 6 Building Analysis as the Basis for Drying Measures (translated from ÖNORM B 3355-1)

The moisture measurements are influenced by various factors, including material properties, the type and concentration of salts, fluctuations in drying conditions and seasonal variations, and ultimately, environmental factors such as relative humidity (Snepvangers, 2005). Snepvangers continues that salt concentration could varies significantly and the concentration can be high in areas where there is no visible damage. When selecting sampling locations it should be considered that salt is not distributed evenly throughout the masonry. Additionally there have been multiple advancements made in the recent years on the study of conditions of crystallisation in relationship with relative humidity and temperature (Charola & Bläuer, 2015).

2.1.5.3. Measurement Techniques

While a variety of laboratory methods exist to identify and quantify salt contamination in masonry—such as electrical conductivity testing, ion chromatography, and photometry (Blauer et al., 2001; Rijksdienst voor de Monumentenzorg, 2005)—these techniques fall outside the practical scope of this research. Instead, this study focuses on the surface-level visual detection of salt damage, particularly efflorescence.

Nonetheless, understanding the underlying salt behavior is important when interpreting visible surface deposits. In situ techniques, such as selective strips that detect specific anions, provide insight into salt composition without full laboratory analysis. However, the results of these methods are highly dependent on the samples themselves, which can vary with seasonal conditions, spatial location, and sampling depth. Salt concentrations may differ significantly by height or depth within the wall, and they can also fluctuate over time (Charola & Bläuer, 2015b). This variability highlights the limitations of visual-only assessment and supports the need for caution when inferring the extent of salt damage from surface efflorescence alone.

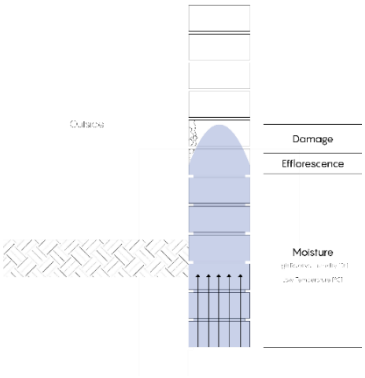
2.1.5.4. Indicators of Efflorescence

The damage originated by soluble salts in porous building materials like masonry are based on different behaviours. Blauer explains that each salt has a different solubility in water, in general there are multiple types of salts present in porous materials where the development of crystallisation pressure is the result of so called “non equilibrium conditions” i.e. explained as the normal conditions found in real life situations (Charola & Bläuer, 2015b).

RISING DAMP : Salt damage in masonry is often linked to rising damp, where groundwater containing dissolved salts is transported into walls due to capillary action as seen in **figure 7 and 8**. This typically occurs in walls without horizontal barriers to block the moisture. As the water evaporates, the concentration of salts increases, eventually leading to crystallization (Lopez-Arce et al., 2009). The salts mainly found oscillate between 15 cm and up to 100 cm but dampness can still be found up to 300 cm though Even after the moisture source is removed, the residual salts left in the walls can crystallize as the wall dries, causing further damage (Charola & Bläuer, 2015b).

Figure 8 Rising Damp (Rafton, 2023)

Figure 7 Rising damp schematic section of a masonry wall



Salt migration and crystallization in masonry is influenced by the solubility of the salts. Less soluble salts tend to crystallize in the lower, more moisture-rich zones of the wall, as they precipitate earlier during the evaporation process. In contrast, highly soluble salts, such as alkali salts, can migrate further upward within the masonry and crystallize at higher elevations. Some salts, like magnesium sulphate, may form through secondary reactions as other salts evaporate. Additionally, salts that do not share a common ion can influence each other's solubility, leading to crystallization in areas that would not be expected based on their individual behavior. De-icing salts, for example, often do not crystallize readily but retain moisture, which sustains damp conditions and intensifies rising damp over time (Charola & Bläuer, 2015b).

LEAKAGE

Efflorescence can also be caused by other moisture sources such as leakage or rain. In the case a leakage is present, efflorescence appears areas, where water infiltration occurs.

In cases where the leakage originates from the roof or gutters, the pattern of efflorescence may follow the path of water run-off along the corners of the building. This creates streaks or patches of efflorescence that appear irregularly along vertical or diagonal lines as seen in **figure 9 and 10**.

The efflorescence extends from the top of the pipe upwards, suggesting leakage from the drainage system. The contrast between the affected and unaffected brick areas emphasizes the localized nature of moisture-related damage.

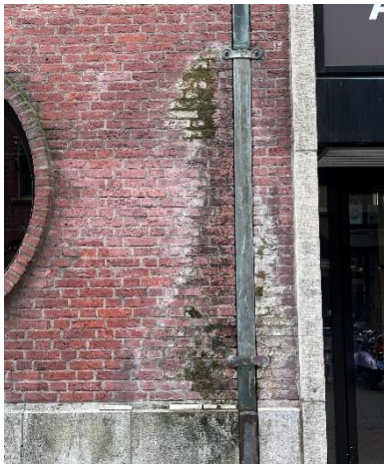


Figure 10 Efflorescence and Biological Growth Near a Drainpipe on a Masonry Wall

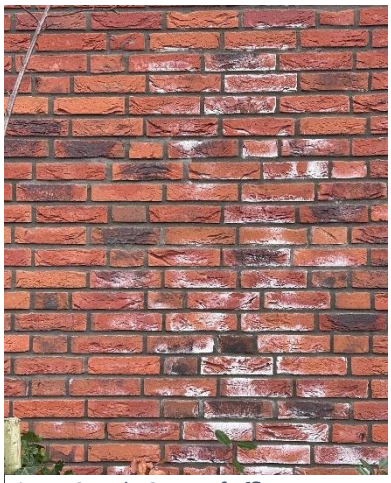


Figure 9 Early Onset of Efflorescence on a Newly Built Masonry Wall

Damage Co-occurrence

Blauer reports that when the flow of water is slower than the evaporation rate of the water, the evaporation front will move towards the inside of the material resulting in the crystallisation within the material and creating damage. There are a number of related damage patterns that can be recognised according to the MDCS and its related research as highlighted in **table 1**:

2.1.5.5. Conclusion & Challenges

Soluble salts leading to efflorescence originates from various sources. Not only by the type, location, and amount of salts present in masonry are influenced by the porous structure of the material, but also moisture amount, distribution and thermohygrometric conditions such as temperature and relative humidity. The pressure exerted by crystalisation in the pores can lead to material degradation. Crystallization of salts at the surface of a material leads to efflorescence, crystallization

of salt in the pores of the material lead to pressures causing damage. As described in the MDCs, salt damage manifests in various forms, such as detachment, loss of adhesion, layering(Vanhellemont, 2008). Rising damp and the hygroscopic properties of materials are key factors contributing to salt accumulation and recrystallization.

Detecting efflorescence presents several challenges. Its appearance may vary depending on the time of day/ year. Furthermore, distinguishing efflorescence from similar phenomena, such as biological growth, or encrustation, i.e. deposit of calcium carbonate deposits can be challenging , based only on images. High salt concentrations may also be present in locations where no immediate visual damage is evident.

In conclusion, salt crystallization is a well-documented and significant factor in the deterioration of historical masonry. Extensive research and observed damage patterns in heritage structures across various geographical regions underscore its critical role in masonry decay.

For the purpose of the machine learning component of this study, a selection of damage types from the Monument Damage Classification System (MDCS) will be incorporated, with the scope tailored to the available dataset and the visual characteristics of the damage. The primary focus will be on damage types categorized under *disintegration*, particularly *loss of cohesion* and *layering*. These types were selected due to their relatively high frequency in the annotated dataset and their relevance to salt-related deterioration processes, such as those triggered by crystallization pressures beneath surface layers.

In addition to disintegration-related forms of damage, the research will consider visually similar *deposit*-type phenomena. These include *graffiti*, *encrustations*, and *lichens*. While graffiti is an anthropogenic surface deposit unrelated to material decay processes, it shares visual features with other white surface deposits and is relevant from a classification perspective. Encrustations, defined in the MDCS as surface accumulations typically resulting from environmental or chemical processes (e.g. calcium carbonate), can resemble efflorescence in texture and tone. Molds, which fall under the broader MDCS category of *biological growth*, are another form of surface deposit that can lead to misclassification due to their light coloration and patchy distribution on masonry. While these deposit types are not damage mechanisms in the structural sense, they pose a risk of false positives during image-based classification and therefore warrant inclusion in the detection framework.

This focused selection allows the model to differentiate between genuinely salt-related damage manifestations and visually similar but unrelated surface conditions, thereby improving classification accuracy and interpretability within the context of efflorescence detection.

2.2. Machine Learning For Damage Detection

This chapter bridges masonry assessment with AI-based solutions. This chapter contains an overview of Machine Learning Models, their applications and architecture.

Structural Health Monitoring (SHM) has long been a critical component of building lifecycle management, with extensive studies conducted on monitoring systems, material degradation, and inspection techniques (Wang et al., 2019). However, over time, the methodologies and technologies used have evolved significantly. In particular, the last decade has witnessed an acceleration in the adoption of machine learning within SHM and damage detection workflows. What was once considered "state-of-the-art" rapidly becomes outdated, as newer algorithms and computational strategies are developed and deployed (Marín-García et al., 2023).

This shift has transformed the fundamental questions guiding damage assessment. Historically, the focus was primarily on classification, determining whether damage was present or not. Today, the emphasis has shifted toward localization and quantification, identifying where damage exists and to what extent, often through region-based segmentation. This transition from image-level classification to pixel-wise instance segmentation has become the new standard of accuracy in the field (Marín-García et al., 2023).

As discussed in the previous chapter, visual inspection methods, especially those performed on-site—often rely on professional judgment and specialized equipment. While these methods remain vital, they are also time-consuming, subject to human error, and increasingly insufficient for large-scale or hard-to-access sites (Hatir et al., 2020; Hatır et al., 2021). With the growing need for efficient, reliable, and scalable assessments, especially in heritage structures with difficult access points or large façades machine learning offers a promising alternative.

The supervised machine learning process for damage detection typically follows a step-by-step pipeline, as shown in **Figure 11:** *General workflow of a supervised machine learning method for damage detection*. It begins with data acquisition and pre-processing, followed by feature extraction to identify relevant patterns. These features are used in statistical modelling and model training, allowing the system to learn from labeled data. Once trained, the model performs damage recognition. If damage is detected, it proceeds to localize and characterize it, ultimately leading to a final damage detection output (Pan et al., 2018).

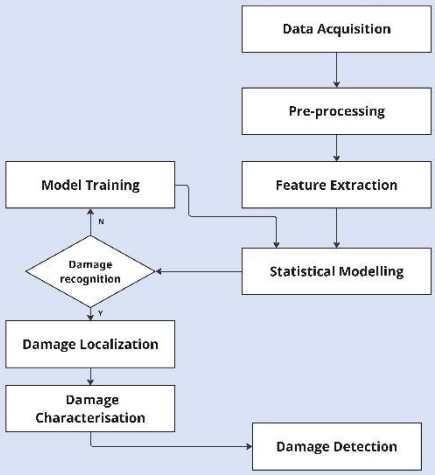


Figure 11 General workflow of a supervised machine learning method for damage detection. Adapted from Pan et al. (2018),

Machine learning not only improves detection speed but also enhances consistency and objectivity. Particularly when combined with tools like drone imaging or point cloud data, Machine learning can process large volumes of visual input and extract meaningful patterns that support conservation, maintenance, and restoration efforts.

Efflorescence often appears with subtle, varied textures and irregular shapes. Its presence can be widespread or localized, and its visibility is sensitive to lighting and surface colour. These characteristics make it especially suitable for machine learning, particularly segmentation models like CNNs that can detect and outline damage with high precision.

2.2.1. Development of Object Detection

The success of machine learning in image classification makes it a promising approach for damage detection. In the past years there have been multiple studies on the detection of damages by means of a convolutional neural network.

Before delving into the application of deep learning in object detection, it is important to acknowledge a persistent ambiguity in the literature regarding the terminology used in computer vision. As noted by Andreopoulos and Tsotsos (2013), terms such as *detection*, *localization*, *recognition*, *classification*, *categorization*, *labeling*, and *understanding* are often used interchangeably or with varying definitions across studies. **Figure 12** illustrates the different computer vision tasks adapted from Lui et al (2020). **Image classification** assigns a single label to an entire image based on its overall content. **Object detection** locates and classifies multiple objects within an image using bounding boxes. **Semantic segmentation** assigns a class label to each pixel, grouping pixels that belong to the same class into one region. **Instance segmentation** further refines semantic segmentation by distinguishing between individual instances of the same class (e.g., multiple windows are individually separated). These distinctions are essential when developing and evaluating machine learning models for damage detection, as each task offers a different level of granularity and interpretability.

Figure 12 Illustration of the primary computer vision tasks: (a) image-level classification, (b) bounding-box object detection, (c) pixel-wise semantic segmentation, and (d) instance-level segmentation. Adapted and illustrated based on Liu et al. (2020).

This lack of standardized vocabulary reflects the evolving and interdisciplinary nature of the field, but it also introduces challenges in framing and comparing research efforts (Liu et al., 2020). To better frame the practical and theoretical challenges involved in developing object detection algorithms, this

study references the taxonomy proposed by Liu et al. (2020), which categorizes the performance requirements and technical hurdles an *ideal detector* must overcome further described in **Figure 13**.

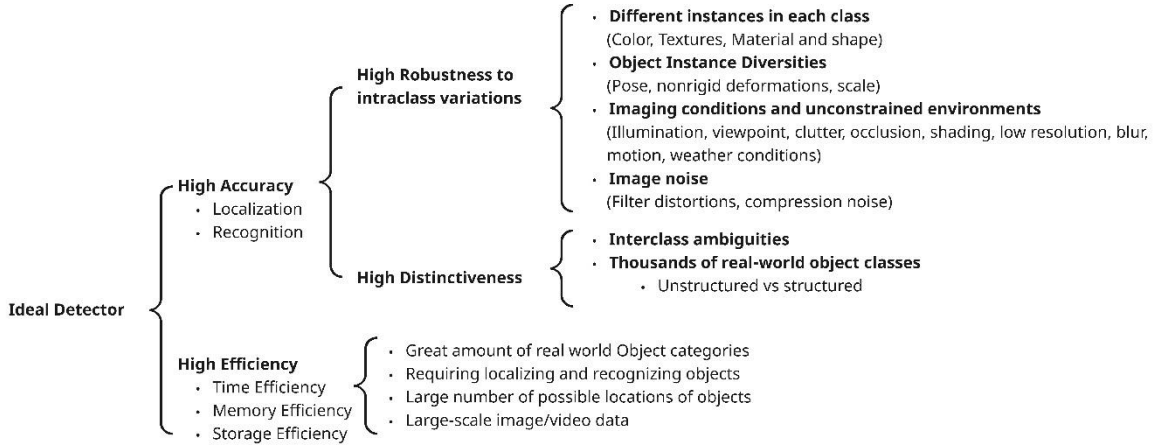


Figure 13 Taxonomy of challenges in generic object detection, outlining the characteristics of an ideal detector in terms of accuracy, efficiency, robustness, and distinctiveness. Adapted from Liu et al. (2020).

In this study, the focus lies specifically on object detection and instance segmentation, which are clearly defined tasks involving the identification of object locations within an image and the delineation of their spatial boundaries, respectively.

In the field of computer vision, particularly object detection, it is important to distinguish between feature extraction backbones and detection frameworks. Backbones such as AlexNet, GoogLeNet, VGG, and ResNet are convolutional neural networks (CNNs) originally developed for image classification tasks, primarily evaluated through the ImageNet Large Scale Visual Recognition Challenge (ILSVRC). These networks are designed to learn hierarchical features from images, progressing from low-level edge detectors to high-level semantic patterns.

In contrast, object detection frameworks such as R-CNN, Fast R-CNN, Faster R-CNN, Mask R-CNN, and YOLO build upon these backbones by integrating additional components like region proposal networks, bounding box regressors, and segmentation heads. The backbone is typically used as the feature extractor, while the detection architecture handles localization, classification, and (in some models) segmentation. The choice of backbone significantly impacts performance and speed, but it is not the detection method itself.

The ImageNet Large Scale Visual Recognition Challenge (ILSVRC) was a benchmark competition that significantly advanced computer vision research. Introduced in 2010, it focused primarily on image classification, with later additions such as object localization and detection. The competition became a proving ground for major CNN architectures, including AlexNet (2012), GoogLeNet (2014), and ResNet (2015). These models demonstrated unprecedented performance on large-scale image classification tasks, marking the beginning of the deep learning era in vision. The development of more powerful and deeper backbone networks has enabled models to learn increasingly abstract and high-resolution feature representations. **Table 2**: *Evolution of key CNN architectures commonly used as backbone networks in object detection frameworks* summarizes key milestones in CNN architecture development.

Table 2 Evolution of key CNN architectures commonly used as backbone networks in object detection frameworks.

Model	Year	Description
AlexNet	2012	Started the deep learning wave with massive improvement
ZFNet	2013	Improved filter visualization and performance
VGGNet	2014	Used very deep (16–19 layer) architecture
GoogLeNet	2014	Introduced inception modules (wider networks)
ResNet	2015	Introduced skip connections for very deep models (up to 152 layers)
DenseNet	2016	Connected each layer to every other layer

By 2015–2017, state-of-the-art models were surpassing human-level performance on the classification task. For example, ResNet (2015) achieved a top-5 error of 3.57%, better than the estimated human error of 5.1%. As models began reaching saturation on classification, research focus shifted toward more complex tasks such as segmentation, pose estimation, and scene understanding. The ILSVRC was officially discontinued after 2017, influenced by a combination of technological maturity, ethical considerations, and the natural progression of research frontiers.

The evolution of machine learning models has shown the growing complexity of damage detection tasks as shown in **Figure 14** *Milestones in generic Object Detection*. Convolutional Neural Networks (CNNs) were the first breakthrough in image classification (LeCun et al., 1998), enabling systems to detect damage categories like cracks or efflorescence in pre-processed photos (Wang et al., 2018). Region-based Convolutional Neural Networks (R-CNN) expanded on this by introducing region proposal mechanisms, allowing the model to not just classify an image but to localize areas of interest within it (Girshick et al., 2016). Fast R-CNN improved efficiency by integrating region proposals and classification into a single, faster network (Ren et al., 2015). Mask R-CNN, the current standard for instance segmentation, introduced a mask branch that performs pixel-level segmentation, allowing not just the detection of damage, but detailed mapping of its shape and spread. Each of these advancements builds upon the limitations of the previous, resulting in higher accuracy, faster processing times, and greater flexibility in dealing with complex geometries and variable conditions.

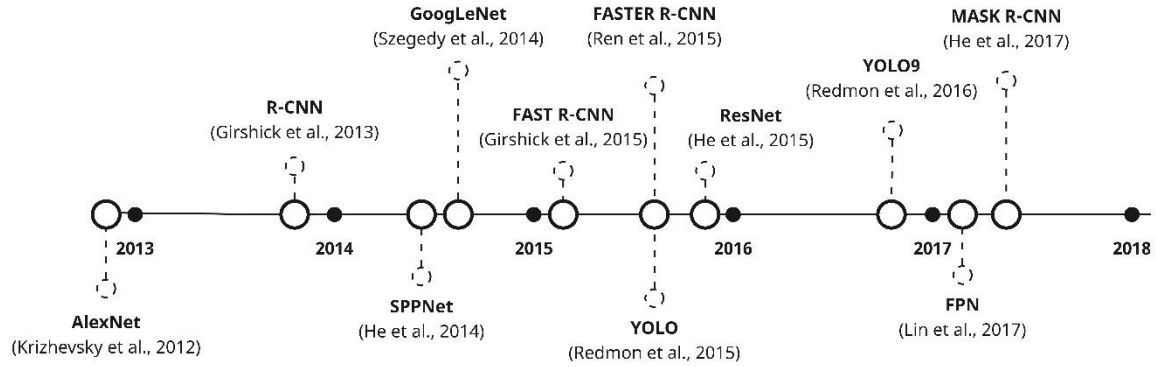


Figure 14 Milestones in generic Object Detection (adapted from Liu et al, 2020)

As object detection models became more capable, research interest moved toward more fine-grained tasks, such as instance segmentation and keypoint detection. In response, the Common Objects in Context (COCO) Challenge was introduced in 2015. Unlike ILSVRC, COCO focuses not just on classifying and localizing objects, but also on pixel-level segmentation, context-aware detection, and multi-object scenes. The COCO dataset contains over 200,000 images with instance-level annotations for more than 80 object categories. It introduced evaluation metrics such as Average Precision (AP)

across different Intersection-over-Union (IoU) thresholds, pushing models to be both precise and robust.

Frameworks like Mask R-CNN were specifically designed with COCO in mind, achieving state-of-the-art results in instance segmentation. COCO continues to be a leading benchmark for evaluating the real-world applicability of object detection and segmentation models.

APPLICATION IN THE FIELD OF HERITAGE MASONRY

The first attempt of classifying and locating multiple types of damages for historic masonry structures based on CNN technique was proposed in 2018 (Wang et al., 2018). Wang continued to achieve a 94.3% accuracy on the classification results for spalling, cracking, efflorescence and intact conditions using the sliding window methodology using the AlexNet and GoogLeNet. Though it must be noted that the dataset consisted of 5145 fixed orthogonal photos of single bricks of the forbidden wall of China (Beijing China). In 2019 an effort was made on the detection of efflorescence and spalling on a small dataset (500 images) consisting of orthogonal homogeneous photo's using Faster R-CNN achieving an mAP of 95% (Wang et al., 2019).

To continue using artificial neural networks an accuracy of 99.4% was obtained on a dataset consisting of 8598 images of orthogonal photo's on eight different classes consisting of fresh rock, flaking, contour scaling, cracking, differential erosion, black crust, efflorescence, higher plants, and graffiti (Hatir et al., 2020). It must be noted that in this case the research limited their scope to the Konya (Turkey) historical site which mainly consisted of Sille stone material. Additionally a research effort was made in Turkey in 2021 on a similar detection of multi class damages (cracks, discontinuities, contour scaling, missing parts, biological colonization, presence of higher plants, de posits, efflorescence, and loss of fresco) using MASK R-CNN. 1740 images were collected from the Gümüsler archaeological site consisting of pyroclastic rocks were the model achieved a mAP of 98.1% (Hatır et al., 2021). Another research effort was made on the detection of efflorescence in Spain by (Marín-García et al., 2023). In their research effort a Yolo v5 (large and small) model was used to train on a dataset of approximately 392 orthogonal images with an mAP of 89.4%. A Multi class detection was used for the bricks classes (1) repair and (2) clear.

In general, the development of datasets and annotation strategies receives limited attention in current research. Mishra et al. (2021) conducted a broad and comprehensive literature review on the application of machine learning in structural health monitoring. While their work provides valuable insights into methodological advancements, it does not address the influence of different annotation strategies on model accuracy. Similarly, the discussion around datasets remains brief, with little consideration given to variations in stone types or material-specific characteristics.

The studies summarised in **table 3** shows the key literature applying machine learning models particularly CNN-based and region-based networks to the task of damage detection in masonry heritage structures. The table highlights datasets, model types, performance metrics, and geographic or material-specific considerations. collectively highlight the evolution of deep learning techniques from basic classification to instance segmentation. As the focus in SHM shifts from detecting damage to understanding it contextually including its extent, shape, and location, models like Mask R-CNN provide the pixel-level granularity needed for high-precision conservation strategies.

Table 3 Overview of State-of-the-Art Machine Learning Models for Damage Detection in Masonry Structures.

Year	Authors	Model Type	Overall Architecture/Used Layers		Types of Damages Identified	Relevance to Efflorescence
2025	Wang F.; Huang J.; Fu Y.	Convolutional Neural Network (CNN)	Two-branch CNN fusing visible, IRT, and microwave data for moisture detection.	Improved accuracy in moisture damage detection through multimodal data fusion.	Moisture-related damage, including water ingress and salt crystallization.	Directly relevant; includes detection of salt crystallization caused by moisture.
2022	Zhou X.; Derome D.; Carmeliet J.	Artificial Neural Network (ANN)	ANN with inputs based on hygrothermal parameters and outputs predicting moisture risk.	Prediction accuracy closely matches simulation results.	Moisture-related damages like condensation, mold growth, and freeze-thaw effects.	Indirectly relevant; identifies moisture risks associated with efflorescence formation.
2021	HatÃ±r M.E.; Ã°nce Ã°.; KorkanÃ§ M.	Mask R-CNN	Region-based convolutional network with mask prediction layers.	Mean Average Precision (mAP): 98.2%, Precision: 91.59â€“100%	Efflorescence, cracks, contour scaling, biological colonization, missing parts, fresco loss, etc.	Directly relevant; efflorescence detection included as a primary damage type.
2019	Wang N.; Zhao X.; Zhao P.; Zhang Y.; Zou Z.; Ou J.	Faster R-CNN with ResNet-101 backbone	Region proposal network, feature extraction with ResNet-101, and bounding box regression layers.	Mean AP: 0.950; Precision for Efflorescence: 0.999.	Efflorescence and spalling.	Directly relevant; includes efflorescence detection as a primary damage category.
2024	Karimi N.; Valibeig N.; Rabiee H.R.	Inception-ResNet-v2	Hybrid architecture combining Inception and ResNet for improved feature extraction.	Accuracy: 96.58%, Precision: 96.96%, Recall: 96.24%	Cracking, flaking, erosion, efflorescence, salt deposition, no defect.	Directly relevant; efflorescence included as a defect type.
2018	Wang N.; Zhao Q.; Li S.; Zhao X.; Zhao P.	Sliding Window CNN with AlexNet and GoogLeNet	AlexNet (8 layers), GoogLeNet (22 layers), employing a sliding window for brick-by-brick analysis.	Achieved an accuracy of 94.3%.	Efflorescence, spalling, cracking, and intact bricks.	Directly relevant; includes efflorescence as a key damage type.
2021	Tijskens A.; Roels S.; Janssen H.	Convolutional Neural Network (CNN)	Custom CNN model optimized for time-series hygrothermal response prediction.	Prediction accuracy: High agreement with simulations (qualitative assessment).	Moisture damage risks (condensation, mold growth, etc.)	Indirectly relevant; identifies moisture risks linked to efflorescence formation.
2024	Alexakis E.; Delegou E.T.; Mavrepis P.; Rifios A.; Kyriazis D.; Moropoulou A.	PSPNet with ResNet-50 backbone	Encoder-decoder architecture with pyramid pooling module for segmentation tasks.	Accuracy: 93%, IoU: 89%, F1-Score: 88%	Rising damp and non-damp areas.	Indirectly relevant; rising damp is often a precursor to efflorescence.
2020	Hatir M.E.; BarstuÄŸan M.; Ã°nce Ã°.	Deep Learning and Artificial Neural Networks (ANNs)	Custom CNN architecture and a fully connected ANN.	DL accuracy: 99.4%, ANN accuracy: 93.95%, Recall: 96â€“100% per class	Efflorescence, cracking, flaking, contour scaling, and others.	Directly relevant; efflorescence is one of the weathering types classified.
2023	MarÃ-n-GarcÃ-a D.; Bienvenido-Huertas D.; Carretero-Ayuso M.J.; Torre S.D.	YOLOv5.	End-to-end CNN for bounding box prediction and damage classification.	mAP: 0.894 at epoch 100, Precision: 89.4%, Recall: 88.6%	Efflorescence (simple cleaning vs. major repair needed).	Directly relevant; focuses on efflorescence classification and repair needs.

Despite their success, current state-of-the-art models still face challenges in generalization across different heritage sites, materials, and environmental conditions. This calls for careful model selection, training strategies, and dataset diversity, which will be discussed in the following sections.

2.2.2. Model Architecture

In recent years, deep learning models have significantly advanced the accuracy and reliability of image-based damage detection as described in the previous chapter. While all models rely on convolutional layers as their foundation, their architecture, purpose, and output vary greatly depending on their design objectives. This chapter compares several major architectures used in visual damage detection workflows with emphasis on their strengths, limitations, and relevance to this study.

It is important to distinguish between the development of application-specific methodologies for object detection (e.g., object detection and segmentation models such as Faster R-CNN or Mask R-CNN) and the evolution of backbone architectures (e.g., AlexNet, VGG, ResNet) on which these methods rely. While methodologies define how an object is detected and represented, the backbone largely determines the feature extraction quality and therefore has a direct impact on accuracy, efficiency, and transferability. For this thesis, understanding both dimensions is essential, since the chosen methodology is closely tied to the capabilities of its backbone..

CNN (CONVOLUTIONAL NEURAL NETWORK)

CNNs are used as an important technique in machine learning and deep learning, specializing in processing grid-like data (images) used for recognition and classification. The CNN architecture as seen in **figure 15** by LeCun et al, (1998) consists of two main parts, (1) Feature extraction and (2) classification. The CNN can be described as a filtering mechanism which goes through different types of filters (layers) to extract features.

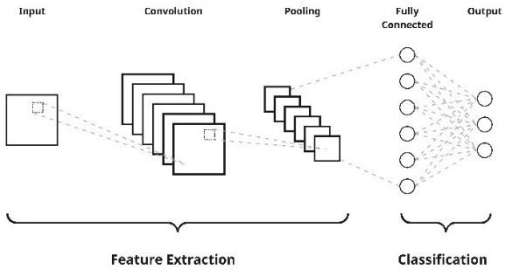


Figure 15 General architecture of a Convolutional Neural Network (CNN), consisting of stacked convolutional, activation, pooling, and fully connected layers (Adapted from LeCun et al., 1998).

The **convolution layers** detects specific patterns such as edges, corners or colour transitions which results in a feature map that highlight the presence of learned features across an image. It could be described as a sliding window that scans the image based on pixel sizes in the form of matrix multiplication.

The **pooling layer** reduces the size of the feature maps to lower the computational load to select the maximum (*MaxPool*) value for each region to preserve the most important features while reducing resolution.

The **Fully Connected Layer** flattens the feature maps into a single vector which is passed through one or more fully connected layers. Each neuron is connected to every neuron in the previous layer such that a prediction can be produced in the form of class scores of the so called SoftMax to output probabilities of each class.

R-CNN (REGION-BASED CONVOLUTIONAL NEURAL NETWORK)

The R-CNN architecture as seen in **figure 16** developed by Girshick et al, (2014) consists of three main parts, (1) Region Proposal (2) Feature extraction and (3) classification. R-CNN used AlexNet as its backbone to extract features from proposed regions. At the time AlexNet had already proven extremely powerful in extracting high level features from images and winning ImageNet 2012 with a significant lead. R-CNN relied on transfer learning which uses a pretrained model for finetuning object detection. Unlike traditional CNNs that process the entire image uniformly, R-CNN first generates region proposals and processes each region individually through a CNN for feature extraction and classification. The downside of R-CNN is by generating +/- 2000 **regions of interest** (ROIs) the model performs relatively slow.

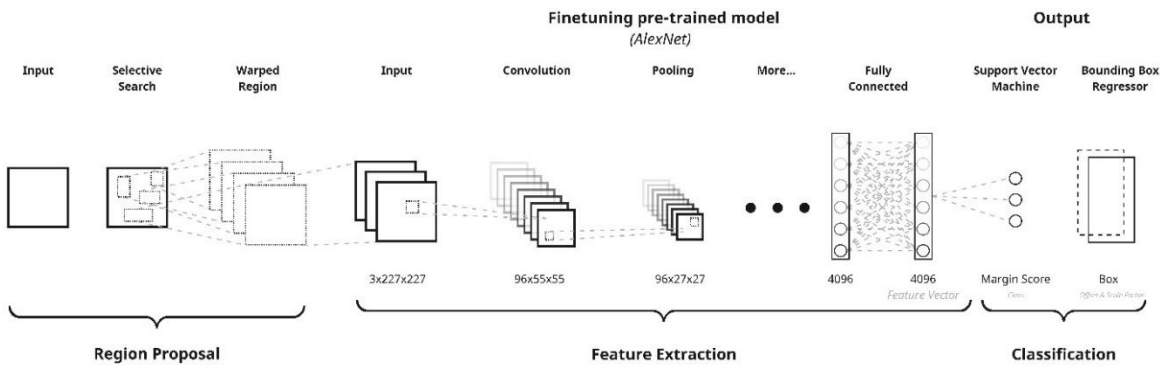


Figure 16 R-CNN architecture: Region proposals are extracted using Selective Search and individually passed through a CNN, followed by SVM classification and bounding box regression (Adapted from Girshick et al., 2014).

The **Activation Function** by means of fully connected layers decides whether a neuron should be activated or not such that in the Fully connected layer determines if it is important in the process of the prediction. The backbone (ie. AlexNet) introduced working with a RelU function to not activate all neurons at the same time in contrast to general sigmoid, Softmax or tanH functions.

It can be questioned whether the success of R-CNN at the time was due more to the underlying deep feature extraction from CNN backbone (AlexNet) and the availability of large datasets like ImageNet, rather than just the novelty of the R-CNN region proposal methodology itself.

FAST R-CNN (FAST REGION CONVOLUTIONAL NEURAL NETWORK)

The Fast R-CNN architecture as described in **figure 17** developed by Girshick (2015) consists of a similar structure as R-CNN with a different Feature extraction. The main difference is the way the region processing through the CNN, namely instead of taking each of the RoI (+/- 2000) and parsing them through the CNN (which requires additional time and computational power), the entire input image is passed through resulting in a shared feature map.

By the same selective search principle as R-CNN, the RoIs are projected onto the shared feature map and then passed through the **RoI pooling** layer resulting in a fixes size feature map (e.g. 7x7) such that a uniform size is maintained. These feature maps are then parsed through the **Fully Connected** layer where then the output is generated with two heads (1) **SoftMax classifier** and (2) **Bounding box regressor**.

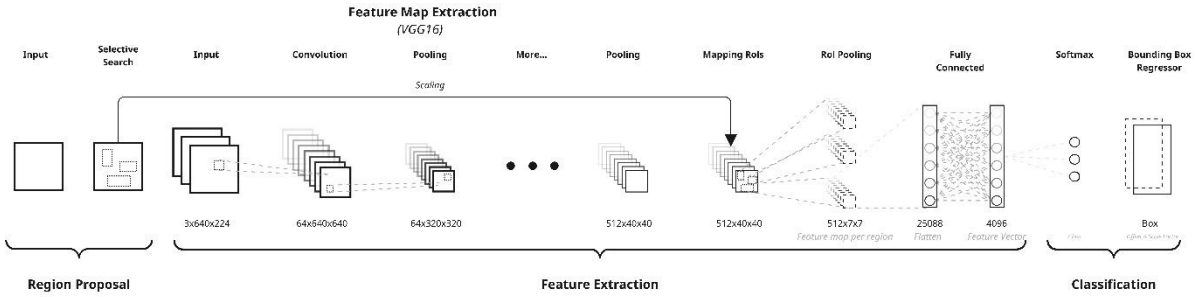


Figure 17 Fast R-CNN architecture: A shared feature map is computed from the full image. RoIs are pooled into fixed-size features using RoI Pooling and processed through fully connected layers for classification and regression (Adapted from Girshick, 2015).

Most interesting improvements are the avoidance of redundant computation through CNN since of the approximately 2000 RoIs only the most “promising” regions are selected at the end. Additionally, the RoI pooling layer preserves the spatial alignment since the “warping” of the image is not applied.

FASTER R-CNN (FASTER REGION CONVOLUTIONAL NEURAL NETWORK)

Similar to the FAST R CNN architecture, FASTER R CNN as described in **figure 18** is also build on it predecessor which additional changes in configuration on layer level and some mathematical concepts. FASTER R-CNN introduced by Ren et al (2015) eliminates the external region proposal algorithm **(Selective Search)** and introduced **Region Proposal Network** (RPN).

In contrast to FAST R-CNN, the RPN "slides" a small network (3x3) by means of a **convolution layer** over the feature map. On the feature map, k-anchors are placed of different sizes per pixel to predict two outputs, (1) **the Objectness score** (background or foreground) and (2) **bounding box coordinates**.

The predicted bounding boxes have different confidence score and might overlap thus might develop problematic predictions, and by means of a **Non-Maximum Suppresion (**NMS) only the most confident boxes remain. Typically a **Intersection over Union** (IoU) threshold of > 0.5 is maintained for this operation.

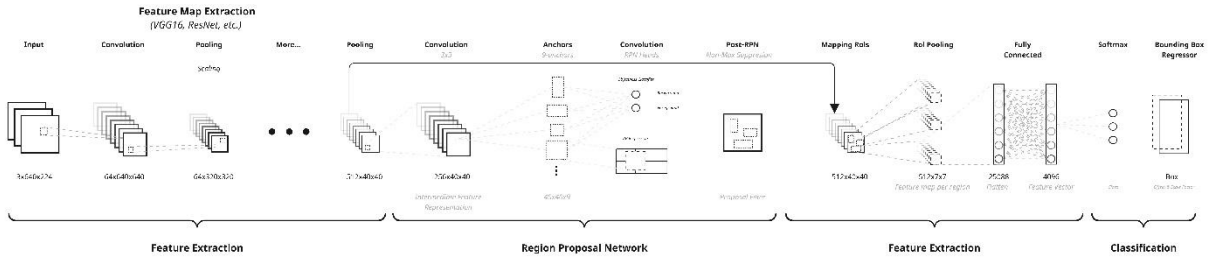


Figure 18 Faster R-CNN architecture: Builds on Fast R-CNN by introducing a Region Proposal Network (RPN) that shares the convolutional backbone and generates region proposals (Adapted from Ren et al., 2015).

Similar to the FAST R-CNN the predicted bounding boxes from the RPN are mapped on the feature maps from the backbone (VGG16, ResNet etc.). These are then parsed to develop feature maps per region which are being handled by the fully connected layers to activate required neurons and develop the classification output with the same heads.

MASK R-CNN (MASK REGION CONVOLUTIONAL NEURAL NETWORK)

MASK R-CNN as described in **figure 19** is an extension of FASTER R-CNN that adds a third branch for predicting the segmentation masks on each RoI in parallel with the existing branches for classification and bounding box regression (wei et al., 2019). Similar to FASTER R-CNN the backbone CNN (ResNet-50 or ResNet-101) extracts features in convolution layers to develop different depths of feature maps.

The Feature Pyramid Network (FPN) is introduced by Lin et al (2017) and applied by He et al, (2017) such that the output feature maps from the backbone CNN (Convolution layer 1, Convolution layer 2, etc) are used to create multi-scale feature maps called P2, P3, P4 etc. The P-levels represent the image feature maps on different scale where P2 is high resolution and good for small objects and P5 is a low resolution good for large objects (Lin et al., 2017).

The integration of FPN Mask R-CNN by He et al, (2017) allows to operate effectively across different object scales, making it a **scalable enhancement** particularly suited for complex scenes with objects of varying sizes. Although not the central innovation, FPN has become a standard component of Mask R-CNN implementations due to its consistent performance gains.

Similar to FASTER R-CNN the feature maps from the FPN are used for the RPN and slides a 3x3 convolution layer across each P-Level. Again for each spatial condition in the feature map

k-anchor boxes are created resulting in a classification and bounding box regression where the top +/- 300 proposals are filtered by means of the NMS.

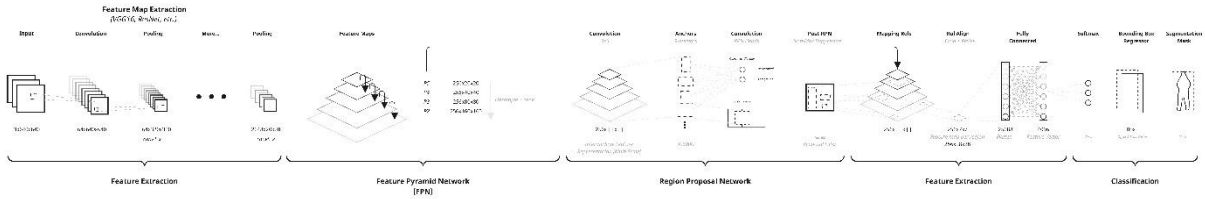


Figure 19 Mask R-CNN architecture with an added mask branch for instance segmentation, RoI Align for pixel-level accuracy, and a Feature Pyramid Network (FPN) for multi-scale feature extraction (Adapted from He et al., 2017).

Instead of **RoI Pooling**, **RoI align** is introduced which instead of using **quantization** or rounding of regions **bilinear interpolation** is used which preserves Pixel-level precision (as required for generating the masks) in the original image.

Finally for each RoI that is mapped on the feature maps from the backbone architecture is parsed through the fully connected layers to develop the three output heads, (1) Classifier, (2) Bounding Box Regressor, (3) Mask Prediction.

YOLO (YOU ONLY LOOK ONCE CONVOLUTIONAL NEURAL NETWORK)

While YOLOv12 by Ultralytics. (Tian. et al, 2025) represents the latest advancement in the YOLO series, this study chose to focus on YOLOv8 due to its broader documentation, extensive community support, and proven performance across diverse benchmarks at the time of research. YOLOv8 has been widely adopted in both academic studies and real-world applications, making it a more stable and interpretable model for comparative analysis. Additionally, the lack of peer-reviewed publications and implementation maturity for YOLOv12 at the time of writing made YOLOv8 a more practical and academically justifiable choice for in-depth exploration and evaluation.

YOLOv8 builds upon the original YOLO architecture introduced by Redmon et al. (2016), represents a one-stage object detection model designed for real-time performance developed by Ultralytics (2023). Unlike two-stage models such as Mask R-CNN, which first generate region proposals and then perform classification and segmentation, YOLOv8 performs all predictions in a single forward pass of the network. The YOLOv8 model consists of three main components, (1) **Feature extraction** also called the *backbone,* (2) **Multi-scale Feature Fusion** (similar to a Feature Pyramid Network) also called the *neck*, and (3) the decoupled **Detection** also called the *Head* which outputs the objectness score, bounding box coordinates and class propabilities directly from the feature maps as shown in **figure 20**.

Feature extraction is similar to the backbone as seen in MASK R-CNN, the main difference is developed in the **Multi-scale Feature Fusion** where the feature maps are developed on multiple scales and processed through an **up sampling** and **concatenation** with lower level feature maps where the **C2f** blocks develop compressed similar output channels (depth). This multi-scale strategy improves accuracy, especially in complex scenes with both small and large objects.

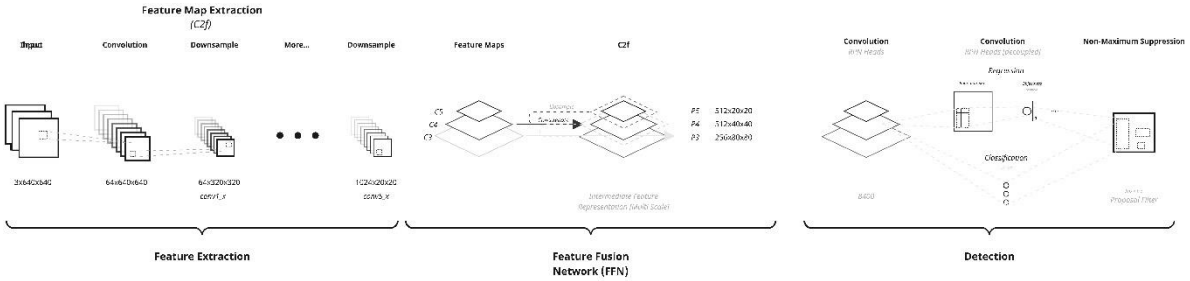


Figure 20 The pipeline consists of feature extraction through a CNN backbone, multi-scale feature fusion via a neck (FPN-like structure), and object prediction through decoupled detection heads. (Adapted from Ultralytics, 2023).

Each spatial location on the feature maps acts as a predictor, estimating whether it contains the centre of an object. YOLOv8 uses a **centre-based**, **anchor-free approach**, avoiding the computational cost and complexity of generating and evaluating thousands of region proposals (Like R-CNN networks). Bounding boxes are predicted relative to each cell, and final predictions are refined using **non-maximum suppression** (NMS).

YOLOv8 achieves its speed by being **fully convolutional and end-to-end**, removing the need for region cropping, fully connected layers, or segmentation masks (unless explicitly added). Predictions are made **directly from the feature maps** without intermediate region-level refinement. As a result, YOLOv8 is significantly faster than Mask R-CNN and more suitable for real-time applications, while Mask R-CNN provides more precise localization and segmentation in tasks where pixel-level accuracy is essential.

YOLOv8 is available in multiple model sizes—**n** (nano), **s** (small), **m** (medium), **l** (large), and **x** (extra-large)—each designed to balance speed and accuracy for different hardware and application requirements. These variants share the same underlying architecture but differ in the number of layers and parameters.

2.2.3. Model selection rationale

While earlier object detection models such as R-CNN, Fast R-CNN, and Faster R-CNN laid the foundation for accurate object localization, their two-stage pipelines and computational demands have been progressively replaced by newer, more efficient architectures (Dupont, 2024). For this reason, this study focuses on two state-of-the-art models that represent the current standard in object detection and segmentation tasks: YOLOv8 and Mask R-CNN.

These models were selected due to their architectural maturity and continued relevance in both research and real-world deployment. A direct comparison between these two models highlights the ongoing trade-off between accuracy and speed, as well as between detection and segmentation granularity.

The performance of both models was evaluated as seen in **Table 4** using key metrics such as **Average Precision** (AP) and **inference speed**, with results reported on the widely adopted COCO dataset.

Table 4 Comparative performance between YOLOv8m-seg and Mask R-CNN on the COCO benchmark. While YOLOv8m-seg achieves significantly faster inference (2–3 ms), its bounding box detection accuracy (AP_{50} = 49.9) is lower than that of Mask R-CNN with a ResNeXt-101-FPN backbone (AP_{50} = 62.3)

Model	AP^{BB} $(COCO)_{50}$	AP^{mask} (COCO)	Segmentation Support	Architecture Type	Inference Speed
Mask R-CNN	62.3	45.8	Yes (pixel-level)	Two-stage	~100+ ms
YOLOv8m-seg	49.9	40.8	No (optional)	One-stage	~2–3 ms

While quantitative comparisons between YOLOv8 and Mask R-CNN provide useful insights, they must be interpreted with caution. Both models are designed for different purposes: YOLOv8 prioritizes real-time object detection, whereas Mask R-CNN is optimized for pixel-level instance segmentation. The **AP^{bb}_{50}** (Average Precision for bounding boxes at 50% IoU threshold) measures how well a model predicts bounding box overlaps, while **AP^{mask}** reflects segmentation accuracy at pixel level. Notably, **Mask R-CNN** tends to excel in segmentation benchmarks due to its two-stage architecture, while **YOLOv8-seg** integrates segmentation as an optional head in a one-stage framework.

Additionally, datasets like COCO span 80 object categories, making results highly dependent on task design, evaluation settings, and model variants. This becomes visible by the comparison research on tree detection of Sapkota (2023) where YOLOv8 had an $mAP_{0.5}$ of 0.902 whereas MASK R CNN performed 0.850 which is significantly less (Sapkota et al., 2023). Though it must be noted that the dataset labellisation methodology seems questionable in this research.

While several models have been reviewed, Mask R-CNN was selected for the following tasks due to its strong performance in instance segmentation and its suitability for detecting complex and fine-grained damage patterns in masonry. However, to contextualize its performance and computational cost, a comparative benchmark with YOLOv8 was conducted further described in chapter 3. This provides a clearer understanding of the trade-offs between segmentation accuracy and inference speed for damage detection applications.

2.3. Key Challenges Identified in the Literature

The literature reviewed in this study highlights the growing role of deep learning in the detection of architectural damage, including efflorescence, across different heritage sites. However, while promising results are presented in various stud ies (e.g., mAP > 0.85 in [Garcia et al., 2023]; precision above 96% in [Kimini et al., 2024]), the reality of applying these models across diverse, real-world environments is considerably more complex. Based on the reviewed work and testing, several critical issues have surfaced that lay the groundwork for deeper investigation.

Misclassification Challenges in Efflorescence Detection

One of the primary difficulties encountered in the detection of efflorescence lies in its visual similarity to other damage types. White deposits on masonry may result from efflorescence, but similar appearances occur due to biological growth (e.g., lichens, algae), encrustations, and even surface graffiti. Previous studies have also reported this limitation, with misclassifications emerging particularly in areas where discoloration and crusted textures were ambiguous. This underscores the need for multi-class damage modelling to distinguish efflorescence more effectively within its visual context.

The Role of Moisture and Infrared Thermal Imaging

Efflorescence is related to moisture transport and evaporation processes. However, as observed in several studies (e.g., Wang et al., 2025), detecting moisture phenomena such as rising damp remains

difficult using only RGB imagery. Moreover, the appearance of efflorescence can fluctuate depending on rainfall, humidity, or evaporation rates. Several studies highlight that thermal infrared (IR) imagery offers strong potential for improving efflorescence detection by capturing moisture-related patterns that RGB alone cannot provide. Wet zones, detected as colder by IR, can align with efflorescence and damage distribution, helping in a more supported identification of efflorescence. Moreover, the combination of efflorescence and moisture distribution could give indications about the most probable moisture source (this last is not the scope of this thesis)

Dataset and Material Limitations

A recurring problem across literature and the dataset relates to variation in building materials and condition. Efflorescence forms differently on bricks, eroded joints, lime mortar, or porous stone surfaces. Papers such as Hatir et al. (2021) and Alexakis et al. (2022) point to the role of material composition and surface degradation in moisture migration, which in turn shapes the development of efflorescence. Yet these factors are often underrepresented in public datasets, leading to poor generalization when models are tested across a range of heritage sites. This issue is compounded when image resolution, angles, or lighting conditions vary, which can obscure subtle white staining or result in loss of surface texture essential for correct classification.

Location, Orientation & Environmental Influence

Further complexities arise in the spatial and environmental context of efflorescence. Like other damage types, also efflorescence is influenced by façade orientation, local weather exposure, and proximity to moisture sources (e.g., sea spray, leaking gutters, or rising damp from ground contact). Some studies highlight distinct efflorescence distributions like horizontal streaking from rising damp, localized spotting from leaks, or even homogeneous salt accumulation internally. Unfortunately, such context is often not modelled explicitly. Without spatial annotations or metadata, models struggle to account for why efflorescence appears in certain areas and not others, thereby reducing prediction reliability.

Need for Expanded and Multi-Layered Annotations

Finally, the literature results support the argument that efflorescence rarely occurs in isolation. In many cases, efflorescence is accompanied by other decay types induced by salt crystallization within the pores, such as loss of cohesion, most often in the form of powdering, scaling, and sometimes spalling. Despite this, most current dataset uses single-label annotation, which cannot represent co-occurring damages. This confirms the need for multi-label training datasets and hierarchical class relationships that can better represent efflorescence and its relationship with other types of decay.

3. Experimental Methodology

This chapter outlines the experimental framework developed to investigate how various contextual and visual factors affect the performance of deep learning models in detecting efflorescence on masonry surfaces. The process begins with the development of a **baseline model**, trained on a curated dataset using two widely adopted object detection architectures: Mask R-CNN and YOLOv8. This benchmark provides a reference point for model performance under standard conditions and highlights initial limitations and challenges in real-world detection scenarios. Additionally, after evaluation of these models, a selection will be made for the model with the highest accuracy ton continue on the hypothesis as stated below.

Building on the insights gained from this baseline, the study then systematically tests a set of targeted hypotheses derived from both literature and practical experience during model development. These hypotheses address factors believed to influence detection accuracy and reliability, including:

- **H1**: The combined distribution of moisture and efflorescence can improve the identification of efflorescence compared to using visual appearance alone.
- **H2**: Efflorescence will be visually misclassified more frequently when other similar looking surface changes (e.g., graffiti, lichens, encrustations) are present in the dataset.
- **H3**: The presence of contextual surface damage (e.g., powdering, scaling) increases the likelihood of efflorescence co-occurring in the same area.
- **H4:** Variations in image quality, angle, and distance negatively affect model performance in detecting efflorescence.

These hypotheses were selected to reflect real-world challenges observed during model testing, as well as insights derived from the literature on material degradation and machine learning-based image analysis. The objective is not only to validate or reject these hypotheses, but also to gain a deeper understanding of the conditions under which detection models like Mask R-CNN and YOLOv8 perform reliably or fail.

Each hypothesis is tested through focused experiments using relevant subsets of the dataset and, where necessary, additional data modalities such as thermal imaging. While the benchmark results are presented in the next chapter (4.1), this chapter explains the methodological choices that support both the baseline and hypothesis-driven evaluations. These include the model training setup, annotation strategies, metric definitions, and analysis procedures.

The goal is to understand not only how well each model performs, but under what circumstances their predictions succeed or fail—an important step toward reliable damage detection in heritage conservation. The following sections provide detailed descriptions of the research approach, data preparation, and evaluation procedures used throughout this study.

The following sections explain the structure and reasoning behind these tests, starting with the design of the overall research approach.

3.1. Research Approach

This research adopts a comparative experimental methodology as described in **figure 21** rooted in machine learning validation. The core objective is to investigate whether specific contextual or visual factors influence the performance of efflorescence detection using deep learning models. To achieve this, a baseline Mask R-CNN model was trained on RGB imagery annotated for efflorescence, serving as the benchmark model against which experimental variants are compared.

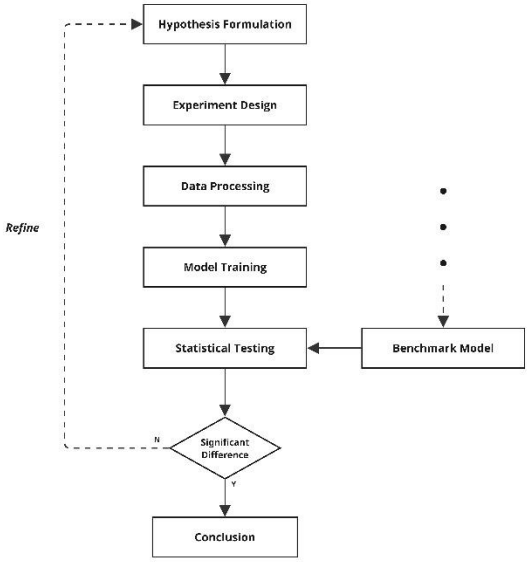


Figure 21 Iterative Experimental Workflow for Hypothesis Testing

Each hypothesis is explored through a controlled evaluation setup, where only one variable is altered at a time (e.g., the addition of thermal imagery, reduced image quality, or presence of similar-looking damages). This structure enables direct comparison of model outputs under different conditions, isolating the influence of each factor.

The approach is both quantitative and qualitative. Metrics such as mAP, precision, and recall are used to measure performance changes numerically, while visual inspection of segmentation and classification outputs supports deeper interpretation. Special emphasis is placed on the multi-modal nature of some experiments, such as the integration of infrared thermal data, to assess their added value in efflorescence detection.

By employing this strategy, the study aims to build not only a performance profile of the baseline model but also to validate or refute the relevance of each contextual hypothesis through measurable and interpretable results.

The performance of deep learning models, particularly in tasks such as damage detection, is highly dependent on the quality and structure of the dataset. Preparing a reliable dataset involves not only collecting relevant images but also ensuring consistent annotations that reflect the specific task. In this project, a custom dataset was developed with a focus on masonry degradation, particularly efflorescence, using both manually annotated images and existing visual inspection data. Careful attention was given to annotation granularity, class balancing, and image resolution, all of which impact model training and generalization.

DATASET REQUIREMENTS

Determining the appropriate dataset size for effective training of a deep learning model is highly dependent on task complexity, class diversity, and model architecture. While there is no universally agreed-upon number of images required for segmentation tasks, previous studies on efflorescence detection provide a helpful benchmark.

For instance, Hatır et al. (2021) used 1,740 images from the Gümüşler archaeological site to train a Mask R-CNN for multi-class damage detection, including efflorescence, and achieved a mean Average Precision (mAP) of 98.1%. In another study, Marín-García et al. (2023) trained a YOLOv5 model on approximately 392 orthogonal images with efflorescence and repair annotations, reporting an mAP of 89.4%. Wang et al. (2019) conducted Faster R-CNN training on a smaller dataset of 500 images, still achieving promising results with a narrower class range.

These studies suggest that even relatively small datasets (ranging from several hundred to a few thousand images) can yield competitive performance, provided the annotations are precise and the data well-structured. Nonetheless, it must be noted that many of these datasets were either focused on highly controlled photographic conditions (e.g., orthogonal images of individual bricks at a fixed distance) or limited to specific material types and environmental contexts, which may affect generalizability.

Despite the growing interest in automated damage detection, there is currently no consensus in the literature regarding the minimum dataset size or required diversity in terms of image conditions (e.g., angle, lighting, and distance). This remains an underexplored yet critical factor in the development of robust, general-purpose detection models.

DATASET DIVERSITY

The dataset developed initially for this research purpose contains a total of 211 verified images with a wide range of resolutions. Image dimensions vary significantly as shown in **Figure 22** *Distribution of image resolutions in the efflorescence dataset*, from small scales such as 232×300 pixels to high-resolution captures up to 5858×3911 pixels. Common dimensions include 1536×2048 pixels (50 images), 4000×3000 pixels (17 images), and 945×709 pixels (20 images), indicating a mix of smartphone and professional camera sources. This diversity in image scale presents both an opportunity and a challenge, requiring resizing or augmentation strategies during preprocessing to ensure consistent model input and effective training.

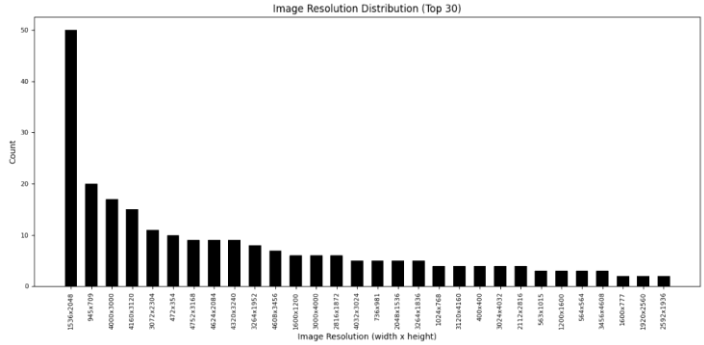


Figure 22 Distribution of image resolutions used in the efflorescence dataset. The graph presents the 30 most frequent image dimensions (in pixels), showing a large variation in resolution across samples.

To enhance the robustness and generalizability of the model, the dataset incorporates a broad range of visual diversity. As shown in *Figure 21*, this includes differences in framing (close-ups versus façade-wide views), surface textures, lighting conditions (natural daylight, low-light environments),

and occlusion from surrounding elements such as vegetation or infrastructure. Additionally, the dataset varies in image quality—ranging from sharp, high-resolution captures to pixelated or motion-blurred images—along with efflorescence intensity, which spans from subtle staining to dense crusts as seen in figure 23. Perspective differences caused by camera angles and the decision to annotate efflorescence on joints or full bricks also contribute to learning complexity. Accounting for these factors is critical, as noted by Liu et al. (2020), who emphasized the role of intraclass variability and environmental conditions in object detection challenges.

Figure 23 Visual examples of dataset diversity in efflorescence images. Variations in framing, surface texture, lighting, occlusion, image quality, intensity, camera angle, and efflorescence location

To improve the model's ability to generalize and reduce overfitting, data augmentation was applied during the training process. Data augmentation artificially increases dataset diversity by introducing variations in the training images, simulating real-world conditions that may not be fully represented in the original dataset. This is particularly important for damage detection tasks, where factors such as lighting, texture, and scale can vary significantly in practice.

In this study, common augmentations included horizontal and vertical flipping, random rotation, brightness and contrast adjustments, and zooming. These methods help the model remain robust when encountering different viewpoints, camera angles, and environmental conditions.

Such strategies are widely recognized in existing research. For example, Bansal et al. (2022) applied brightness, scaling, and rotation to improve efflorescence classification on masonry surfaces, while Saleh et al. (2021) demonstrated the benefits of flipping and cropping in improving model performance on salt crystallization datasets. Similarly, Zhou et al. (2020) highlighted how image distortion and contrast variation improved deep learning models in wall defect detection tasks.

A recurring challenge in constructing a high-quality dataset for heritage-related damage detection lies in the accessibility and ownership of expert-labeled data. Despite concerted efforts to expand the dataset through external resources, limitations in data sharing posed significant barriers. Platforms such as the *Masonry Damage Diagnostics* initiative, which host valuable annotated images of masonry pathologies, were not accessible due to licensing restrictions and institutional data governance policies.

Moreover, attempts to obtain data through direct contact with multiple research groups and institutions yielded limited results. In most cases, responses were either not received or explicitly stated that data could not be shared due to confidentiality agreements or lack of a clear data-sharing framework. These limitations are particularly pronounced in heritage contexts, where image data is often collected under strict project agreements or governmental oversight, making redistribution complex. To account for these limitations field research was required and the general methodology in this research is revised.

The lack of openly accessible, high-quality annotated datasets in this domain remains a bottleneck for developing and benchmarking machine learning models. It underscores the need for more collaborative and standardized data-sharing efforts within the architectural conservation and heritage science communities.

ANNOTATION STRATEGY

To ensure consistency, accuracy, and compatibility with modern deep learning frameworks, the annotation process in this study was conducted using Roboflow, a browser-based tool widely adopted for its intuitive interface and versatile export functionalities. Roboflow supports polygon-based annotations and enables direct export to popular formats such as COCO and YOLO, making it particularly suitable for tasks requiring both instance segmentation and object detection.

Another benefit of using Roboflow was its built-in capability to train models directly within the platform. This eliminated the need for additional coding overhead and enabled a comparative benchmark between Mask R-CNN and YOLO-based architectures using the same annotated dataset.

For this project, annotations were exported in the COCO format, which is required for training Mask R-CNN models. The COCO (Common Objects in Context) format structures image annotations in JSON files that include segmentation polygons, bounding boxes, class IDs, and image metadata. This structure supports both object detection and instance segmentation tasks, making it an ideal choice for evaluating different model architectures.

Initially, the annotation strategy focused on directly labeling only the efflorescence deposits using class-agnostic, pixel-wise polygon masks. This approach aimed to train the model to identify the presence of efflorescence without distinguishing where it occurred on the masonry surface (e.g., brick face or mortar joint).

However, this approach yielded limited results in early experiments. The subtle visual patterns of efflorescence, especially when it appeared in low contrast or small patches, proved difficult for the model to learn robustly. Moreover, focusing solely on efflorescence without contextualizing its location on individual bricks made it challenging to draw meaningful conclusions about its spatial distribution or potential causes.

As a result, the annotation strategy was revised to a brick-level annotation approach. Instead of labeling just the efflorescence, bricks affected by efflorescence were annotated as entire objects. This

shift aligned better with the overall research scope, which focuses on recognizing and analyzing efflorescence patterns at the building element level, rather than detecting isolated deposits.

An additional attempt was made to implement a multi-class annotation strategy, distinguishing between *efflorescence on brick* and *efflorescence on mortar joints*. This distinction aimed to support more nuanced analyses, such as identifying material-based susceptibility or the influence of joint permeability. However, this approach quickly revealed a significant class imbalance: the majority of images featured efflorescence primarily on bricks, while examples of efflorescence localized on joints were relatively scarce. The resulting data sparsity in the 'joint' class negatively impacted model training and led to unstable performance across categories. Due to this imbalance and the limited benefit for the core research objective, the multi-class approach was abandoned in favor of a single-class annotation focused on bricks affected by efflorescence.

3.2. Experimental Design

This section outlines how each hypothesis was translated into a structured and testable experimental setup. The overall design follows a modular approach, where each hypothesis is examined by introducing a single, controlled modification to the baseline model.

To ensure reliability and isolate effects, all other model parameters, training settings, and evaluation procedures are held constant. Performance is assessed using the same test set and evaluation metrics across experiments, including mAP@0.5, precision, and recall. By maintaining consistency in evaluation, performance differences can be attributed to the specific condition under investigation.

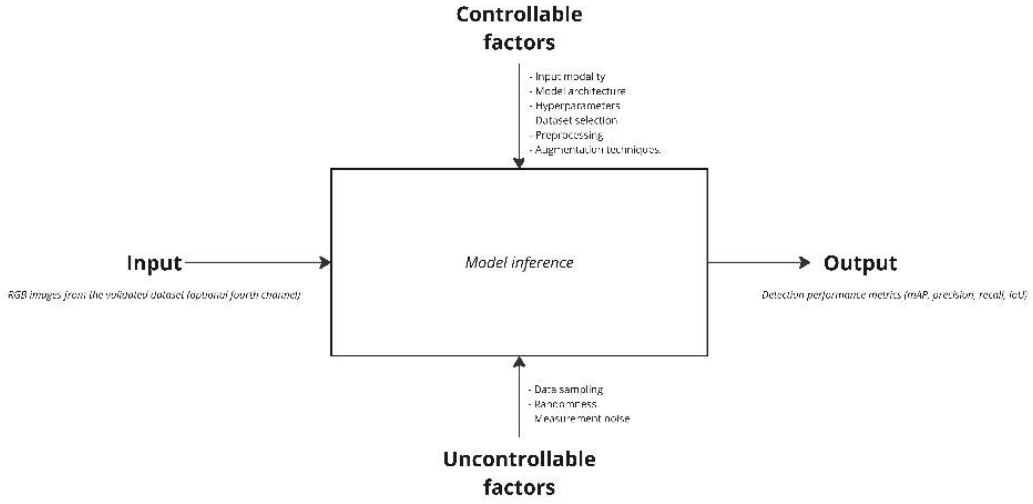


Figure 24 Factors Affecting Deep Learning Model Inference adapted from Kogan (2016)

Maintaining this structure is essential when working with machine learning models, where minor changes in configuration or data can have significant effects on performance. To address this, controllable factors as described in **Figure 24:** *Factors Affecting Deep Learning Model Inference* such as input modality, dataset composition, model configuration, and preprocessing steps are explicitly managed across experiments. In contrast, uncontrollable factors including randomness in training initialization, image sampling variability, or noise in measurements are accounted for by repeated runs, fixed seeds, and consistent evaluation. Each hypothesis in this study translates into a specific experimental condition, outlined below.

H1 Thermal Imaging and Moisture Detection

This hypothesis introduces a fourth input channel derived from infrared thermal images, which are aligned pixel-wise with their RGB counterparts. The thermal signal represents surface temperature variations, with a specific focus on detecting colder areas indicative of moisture presence. This experiment includes 200 annotated image pairs (RGB + thermal) with Efflorescence and Damage regions labelled identically in both modalities by introducing Thermal in grayscale as the 4th input channel. These images cover varied environmental contexts, with an average temperature range of 6–25°C and visible thermal gradients of at least 3°C between dry and moist regions. By integrating the thermal channel during both training and inference, the model's sensitivity to moisture-associated damage patterns is evaluated.

In order to enhance the detection accuracy of efflorescence, thermal imaging data was incorporated as an additional input channel to the model. This required a significant adaptation to the model architecture since the original convolutional layers were designed for three-channel (RGB) input. The first convolutional layer in the Mask R-CNN model is inherently designed for three-channel (RGB) input. To accommodate the four-channel (RGB + Thermal) data, the Conv1 layer was excluded during weight loading, allowing it to be randomly initialized to match the new input shape. The model expects a mean pixel value for each input channel. Since the thermal data is added as a fourth channel, the mean pixel value for this channel was set to 0.0 as a placeholder. This ensured consistency in input preprocessing without distorting the existing model architecture. The thermal images were loaded as grayscale and resized to match the dimensions of the corresponding RGB images as described in the MASK-RCNN repo.

H2 Misclassification Risk Due to Visual Similarity

To examine the risk of misclassification (H2), the experiment evaluates how well the baseline model (trained only to detect efflorescence) performs when presented with images of visually similar but fundamentally different conditions. These include **graffiti, encrustation, biological growth** (e.g., lichens, algae). The initial phase tests the baseline model on a curated set of approximately 150 images representing each of these conditions. These images might not contain efflorescence, and any detections are treated as false positives, quantifying the model's confusion under real-world visual ambiguity.

Following this baseline assessment, a retraining phase is introduced to reduce misclassification. The model is extended to a multi-class setup. Two training strategies are proposed:

- **Per-Class Retraining**: The model is retrained separately for each potential source of confusion. For example, in one experiment, only graffiti and efflorescence are used. This isolates how well the model distinguishes between specific pairs of similar-looking classes.
- **Combined Multi-Class Training**: All new classes are introduced in a single extended training run, creating a comprehensive five-class model. This reflects real-world deployment but increases the complexity of class separation, which may impact performance due to overlapping visual features.

The same evaluation protocol is used across both strategies, comparing metrics such as per-class precision, recall, and confusion matrix-derived false positives. This setup aims to determine whether fine-tuning on more diverse classes helps the model learn subtle visual differences or whether added class complexity degrades overall precision.

H3 Damage Co-Occurrence

H3 focuses on the spatial relationship between efflorescence and adjacent forms of surface loss of cohesion, including powdering, scaling, and spalling. A total of 120 images containing efflorescence co-located with one or more of these damage types are annotated with bounding boxes for both

efflorescence and the secondary damage. The new class *[Damage]* including scaling, powdering, and spalling are included in both training and testing. The experiment tests whether the model's accuracy in detecting efflorescence improves when it learns from contextual features associated with moisture migration and salt accumulation pathways.

STATISTICAL RELATIONSHIP

To statistically evaluate the spatial relationship between efflorescence and other forms of masonry damage, there are multiple options.

The first option is Conditional Random Field, which is limited to a dual class set up, by which the positives spatial test anchored to efflorescence compares distance distribution. A permutation procedure (10,000 label shuffles) generated the null distribution of mean nearest-neighbour distance; the empirical p-value quantified whether damage occurs closer to efflorescence than expected under independence. Limitations of this methodology are the inhomogeneous spatial structure due to objects, edge effects or truncated neighbourhoods from boundaries, pooling of points per image can ignore within image correlation which results in optimistic p-values, double counting due to double classes being detected might cause distance = 0 which can interfere with the neighbourhood signalling. Additionally the damage processes might be directional like rising damp or leakage, this might cause limitations to not use isotropic Euclidean distances.

The second methodology relies on contingency tables and chi-square testing to evaluate whether efflorescence and damage co-occur more frequently than expected under independence. In this approach, the analysis is anchored to efflorescence annotations, and the surrounding area is divided into concentric neighbourhood zones (e.g., within one brick's distance, within two bricks, etc.). For each zone, counts of bricks with and without damage are tabulated, forming an observed contingency table. The chi-square statistic is then used to compare these observed frequencies against the expected frequencies under the null hypothesis that damage occurrence is independent of proximity to efflorescence. An empirical p-value indicates whether damage is disproportionately clustered near efflorescence compared to farther away.

This method is straightforward to implement and directly interpretable, as it quantifies how the probability of damage changes with distance from efflorescence. Moreover, the contingency table framework allows stratification by zone, enabling comparisons of "near," "intermediate," and "far" relationships. However, several limitations apply. First, the results depend heavily on how zones are defined (e.g., brick size, average bounding box width), which introduces subjectivity. Second, imbalance in the number of annotated bricks per image may bias the results, particularly if some images contain many annotations while others contain very few. Third, the chi-square test assumes independence of observations, yet bricks within the same wall segment may not be independent due to shared exposure or construction context. Despite these limitations, this approach provides a statistically grounded way to quantify neighbourhood co-occurrence, and its results can be used to inform confidence calibration of model predictions, either via fixed rule-based adjustments per zone or through regression-based probability recalibration.

Compared to more complex randomization approaches, the chi-square method has fewer limitations. It avoids issues such as edge effects, truncated neighbourhoods, and artificial dependence introduced by pooling across images. Because the chi-square test works directly with observed and expected counts, it is less sensitive to geometric assumptions (e.g., distance distributions) and more transparent to interpret. While it still depends on zone definitions and assumes independence between observations, these constraints are more manageable and easier to justify within the scope of masonry wall analysis.

CHI-SQUARE TEST

A chi-square test of independence (χ^2) was applied. This test assesses whether two categorical variables are independent, or whether there is a significant association between them. In this study, the two variables were: (1) Presence of efflorescence (per zone), (2) Presence of damage (present vs. absent)

A contingency table was constructed by counting how often damaged bricks occurred in predefined distance zones relative to efflorescence (see Section 4.2.3). The expected frequencies were calculated under the null hypothesis that efflorescence and damage occur independently of each other.

The test statistic is defined as:

$$\chi 2 = \sum_{i,j} \frac{(O_{ij} - E_{ij})^2}{E_{ij}}$$

$O_{ij} = observed\ frequency\ in\ cell\ i,j$

Where the observed frequency is defined through the contingency table where the amount of damaged vs undamaged bricks are counted:

ZONE	**DAMAGED**	**UNDAMAGED**	**TOTAL** R_i
Zone 1	O_{11}	O_{12}	R_1
Zone 2	O_{21}	O_{22}	R_2
Zone 3	O_{31}	O_{32}	R_3
COLUMN TOTALS C_j	C_1	C_2	N

$E_{ij} = expected\ frequency\ in\ cell\ i,j$

$$E_{ij} = \frac{R_i \times C_j}{N}$$

$C_j = Column\ Totals{:}\ total\ bricks\ in\ category\ j\ across\ all\ zones.$

$R_i = Row\ Totals{:}\ total\ bricks\ in\ zone\ i$

$N = Grand\ Total$

Then similar to the contingency table, the expected value $[E_{ij}]$ table can be constructed. All contribution can be summarised with respect to $\chi 2$

The degrees of freedom can be calculated by:

$$df = (r-1)(c-1)$$

$r = Number\ of\ zones$

$c = Number\ of\ categories$

Since the chi-square distribution with $df = 2$ has a probability density function such that:

$$f(x;2) = \frac{1}{2}e^{-x/2}, x \geq 0$$

Where

$$p = P(\chi 2 \geq x \mid df = 2)$$

Where the cumulative distribution function where $df = 2$ for $\chi 2$ is:

$$F(x;2) = 1 - e^{-x/2}$$

And to finalize the p-value

$$p = F(\chi 2;2) = 1 - e^{-x/2}$$

Additionally the effect size by (Cramér's V) can be calculated:

$$V = \sqrt{\frac{\chi 2}{N \times (k-1)}}$$

$k = min(c\ or\ r)$

Where Interpretation (Cohen's rule of thumb for 2 categories):

- 0.1 = small
- 0.3 = medium
- 0.5 = large

At last the descriptive percentages of damaged vs undamaged bricks per zone are reported.

CONFIDENCE ADJUSTMENT

With the end goal in mind, the results of this test can support confidence adjustment in detection models, either through a fixed rule-based system that modifies prediction scores per zone, or a calibrated approach based on a logit transformation. In the latter case, the chi-square-derived relationships between efflorescence and nearby damage can be incorporated into a logistic regression model, with coefficients converted back into adjusted probabilities. This allows the raw model outputs to be re-weighted in line with empirical evidence of spatial co-occurrence, thereby improving the interpretability and reliability of automated efflorescence detection.

To refine the confidence of efflorescence detections based on the presence of nearby damage, a logistic calibration model was applied. Logistic regression is a probabilistic model that estimates the likelihood of a binary outcome (here: efflorescence detection being correct) as a function of one or more predictor variables. In this study, the predictors are:

1. Baseline model confidence for efflorescence (p, as predicted by Mask R-CNN).
2. Proximity zone of damage relative to the efflorescence detection (Zone 1 = within 1 brick width, Zone 2 = within 2 brick widths, etc.).

i	BASE PROB $\hat{p}_i$	ZONE Z_i	LABEL y_i
1	$\hat{p}_1$	Z_i	y_i
2	$\hat{p}_2$	Z_i	y_i
Etc.	$\hat{p}_i$	Z_i	y_i

$\hat{p}_i = The\ probability\ per\ damage\ prediction$

$Z_i = The\ specified\ zone\ (1{,}2\ etc)\ set\ predicition\ is\ classified$

$y_i = The\ label\ [0{,}1]\ (damage\ or\ not)$

After the zone predictions are tabulated the base logit from the model's probability can be calculated.

$$\ell_i = \log \frac{\hat{p}_i}{1 - \hat{p}_i}$$

The end goal is to create the Design Matrix X with ℓ, $z1, z2$ with target vector y. Where $z1, = 1(Z = 1)$ are the intercept.

i	BASE LOGIT ℓ,	ZONE $z1$	ZONE $z2$	LABEL y_i
1	ℓ_1	[0.1]	[0.1]	[0.1]
2	ℓ_2	[0.1]	[0.1]	[0.1]
Etc.	ℓ_i	[0.1]	[0.1]	[0.1]

Afterwards the logistic regression model can be developed. In reference to the intercept, slope and zone adjustments.

$$p_i = \sigma(\eta_i), \qquad \eta_i = \beta_0 + \beta_1 l_i + \gamma_1 z_{1i} + \gamma_2 z_{2i}$$

β_0 =Intercept

β_1 = Slope for base

γ_1, γ_2,= $Zone\ adjustments$

Then the Newton-Raphson algorithm can be initialized by:

$$\theta^{(0)} = (0,0,0,0)^T$$

Such that

$$\eta^{(0)} = X\theta^{(0)} = 0$$

And the general updated formula:

$$\theta^{(t+1)} = (X^\top W X)^{-1} X^\top W z$$

If no damage is detected in an image at inference: assign $Z = 3$ (no boost).

If damage predictions are noisy, make the proximity feature robust:

- Use only damage detections with $\widehat{q_j} \geq \tau$
- Cap the maximum boost so a single damage doesn't over-inflate confidence $p' \leq p_{\max}$

Keep the zone definitions identical to those used in the chi-square analysis (same normalization, same thresholds).

H4 Influence of Image Acquisition Conditions

H4 evaluates the influence of image acquisition conditions. A set of 180 test images is used, captured under systematically varied resolution, angle, and lighting.

Camera angle is emulated through rotation augmentations applied in steps of 15°, covering a range from -30° to +30°, resulting in seven distinct orientations per image. Camera distance is simulated via scaling transformations, applied in five steps with zoom factors of 0.8x, 0.9x, 1.0x (original), 1.2x, and 1.5x, mimicking variations in image proximity to the surface. Lighting conditions are approximated using photometric distortions, where brightness and contrast are adjusted in increments of 10% across five levels: -20%, -10%, 0%, +10%, and +20%. This results in a systematically augmented dataset where the influence of each variable can be isolated and assessed.

These conditions are kept consistent across scenes by augmenting images at fixed intensity intervals. The model is retrained using these augmented images and compared against the baseline to assess whether robustness to these conditions can be improved. Evaluation is performed on a held-out augmented test set, and detection performance is reported per augmentation type to identify which conditions most affect model reliability.

3.3. Data Collection

Each hypothesis in this study required a specific dataset, either created through fieldwork or curated from external sources. This section outlines where and how these datasets were collected or constructed.

H1: Thermal Imaging and Moisture Detection

Thermal and RGB images were collected at sites with known moisture-related efflorescence. The goal was to align thermal (infrared) data with visible surface damage. Sites included as seen in figure 25.

Figure 25 Data collection map

At each location, thermal and RGB images were captured under similar framing to allow for manual alignment and annotation. The final dataset aims to include at least 200 aligned RGB and thermal image pairs with annotated efflorescence regions based on the RGB input.

Data was collected using a FLIR T4xx series thermal imaging camera, which captures both infrared and RGB images. The thermal images have a resolution of 320 × 240 pixels, while RGB photos were captured at 1280 × 960 or higher, using a standard DSLR for colour.

Quality control was applied to exclude poorly aligned or noisy thermal captures. Images were removed from the dataset if they showed:

- Saturation artifacts in the thermal channel (e.g., overexposed reflections)
- Alignment errors exceeding 15 pixels across key structural features
- Lighting inconsistencies in RGB captures that compromised annotation accuracy

The final dataset consists of approximately 200 image pairs, each with pixel-level annotations derived from the visible RGB image, applied to the aligned thermal-RGB composite for model training and validation.

H2: Misclassification Risk Due to Visual Similarity

Data sources included a combination of public datasets and original field photography. Public images were screened for resolution, relevance, and visual clarity. Custom images were collected at the following locations, targeting masonry surfaces affected by these misleading conditions.

All images were captured using the same camera setup as the benchmark model (1280×960 resolution RGB). Each damage type was annotated as a separate class using polygon-based masks.

Images with multiple types of damage were included when appropriate, but care was taken to maintain visual distinction between the categories. Images were filtered to remove poorly lit, blurry, or ambiguous samples. Final dataset size:

- Graffiti: 350 images
- Encrustation: 280 images
- Biological growth: 120 images

Due to the limited availability of graffiti and lichen images specifically on masonry surfaces, additional data augmentation was necessary. Graffiti and lichen images were sourced from publicly available instance-segmented datasets. The lichen dataset was obtained from Rojas (2019), while the masonry backgrounds were derived from a large public dataset of 15,000 images (*public dataset, 2024*). To enhance the model's ability to generalize to masonry contexts, the graffiti and lichen images were composited onto the masonry backgrounds using techniques such as Gaussian blur and edge feathering to maintain visual consistency. This approach ensured the creation of a targeted training set that better represents graffiti and lichens on brick or stone surfaces.

H3: Damage Co-Occurrence

The dataset used for this hypothesis was primarily sourced from annotated damage imagery provided by one of the supervisors of this research work, Barbara Lubelli, and supplemented with additional images collected in the field. All relevant damage types were merged under a single class label, damage, to simplify the training process due to dataset size constraints. Sources and Locations:

Supervisor dataset: High-resolution annotated masonry damage dataset, manually verified and segmented, containing 180 relevant images with combined efflorescence and disintegration.

Field-collected imagery:

All images were captured using the standard RGB camera at 1280×960 resolution. A set of 100 additional field images were added to the dataset, and annotations were manually drawn using polygon masks to mark visible disintegration. Images were included if both efflorescence and surface disintegration occurred within the same image frame, even if they affected separate bricks. This approach supports the co-occurrence analysis by allowing the model to learn contextual visual cues that may indicate efflorescence likelihood indirectly.

H4: Influence of Image Acquisition Conditions

To ensure consistent and reproducible testing across the same benchmark test set, augmentations were applied using the Python-based libraries Albumentations and OpenCV. These libraries were selected for their flexibility, reproducibility, and their ability to apply compound geometric and photometric transformations while preserving annotation alignment.

Camera Angle Variation
Rotational transformations were used to simulate changes in camera angle. Each image was rotated around its center at 5°, 10° and 15° (both clockwise and counterclockwise), resulting in 8 rotated variants per image as shown in figure 26.

Figure 26 Camera Angle (Rotation). Left to right: images rotated in steps of 5°, 10°, and 15° relative to the original orientation.

Camera Distance (Scale)

Simulated using zoom in/out scaling as shown in figure 27. Images were resized with zoom factors of 0.5×, 0.0.75×, 1.0×, 1.25×, and 1.5×, creating 5 scale variants per image. Resized images were cropped or padded to maintain consistent input size.

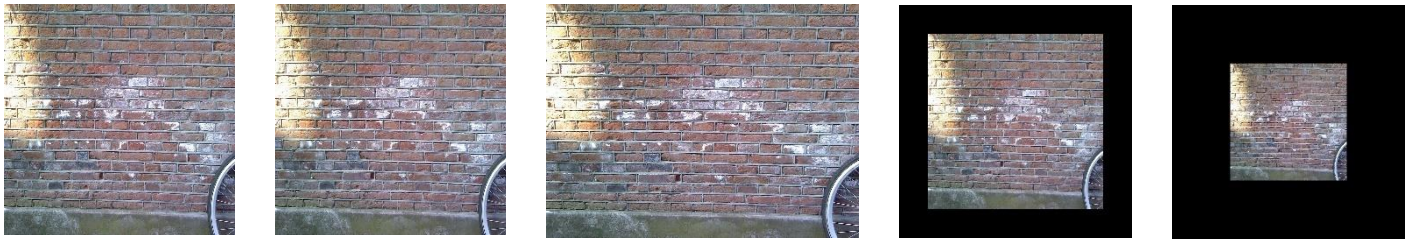


Figure 27 Camera Distance (Scale). Left to right: scaled to 0.5×, 0.75×, 1.25×, and 1.5×, with padding to retain 640×640 resolution

Image Resolution

Downsampling was performed to simulate lower-resolution imagery as shown in figure 28 and re-upscaled to original input dimensions (640×640). Steps included reductions to 480×480, 320×320, and 160×160 followed by bicubic interpolation back to full size.

Figure 28 Image Resolution. Left to right: downsampled to 480×480, 320×320, and 160×160, then upscaled back to 640×640.

Lighting and Exposure

Photometric distortion was applied through controlled contrast and brightness shifts as shown in figure 29. Brightness was adjusted in steps of ±10%, ±20%, and ±30%, and contrast in ±10%, ±20% and ±30%, resulting in 7 lighting variants per image. These changes reflect typical environmental variability (e.g., overcast vs. direct sun).

Figure 29 Lighting / Exposure. Left to right: brightness and contrast adjusted in steps of ±10%, ±20%, and ±30%.

All augmentations were applied to the validation subset of the benchmark RGB dataset. Ground truth masks were transformed alongside the images to ensure perfect alignment. The total number of augmented variants per original image ranged from 7 to 15, depending on augmentation combinations.

3.4. Model evaluation

The model performance is evaluated using both quantitative metrics and qualitative image inspection, aligned with the requirements of multi-class semantic segmentation. For quantitative analysis, three primary metrics are used consistently across all experiments .

The evaluation of the trained Mask R-CNN model is necessary for understanding its effectiveness in detecting efflorescence in masonry surfaces. To objectively assess the model's performance, a combination of standard evaluation metrics, validation loss tracking, and visual inspection of predictions was used. Additionally, evaluation metrics included Precision, Recall, Intersection over Union (IoU), mean Average Precision at IoU 0.5 (mAP@0.5). The mathematical formulations for these metrics, as applied in recent comparative studies such as the work by (Sapkota et al., 2024).

Precision measures the proportion of correctly predicted positive instances out of all predicted positive instances:

$$Precision = \frac{TP}{(FP + TP\,)}$$

Recall quantifies how many actual positives were correctly identified:

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

IoU (Intersection over Union) compares predicted and ground truth masks or boxes by the overlap area divided by the union area:

$$IoU = \frac{Area\ of\ Overlap}{Area\ of\ Union}$$

mAP@0.5 (mean Average Precision at 0.5 IoU threshold) averages the AP across classes using a threshold of 50% overlap between predicted and ground truth masks:

$$\text{mAP}_{@0.5} = \frac{1}{\mathrm{K}}\sum_{\mathrm{i}=1}^{\mathrm{K}} \mathrm{AP_i}$$

Moreover, these evaluation metrics are not only widely adopted in general object detection tasks but have also been consistently applied across recent research on moisture-induced and salt-related deterioration in heritage buildings. For example, in studies listed in the comparative summary table (e.g., Hatir et al., 2021; Kim et al., 2023; Nan et al., 2023), metrics like mAP, precision, and recall were used to assess model performance in detecting damages.

In addition to these standard metrics, confusion matrices are introduced as a new analytical tool in this chapter. While Chapter 2.2.6 focused on binary classification, the shift to multi-class testing (e.g., distinguishing efflorescence from encrustation, graffiti, or biological growth) necessitates deeper insight into inter-class confusion.

Table 5 Confusion matrix for multi-class classification performance (Adapted from Cowan, 2024)

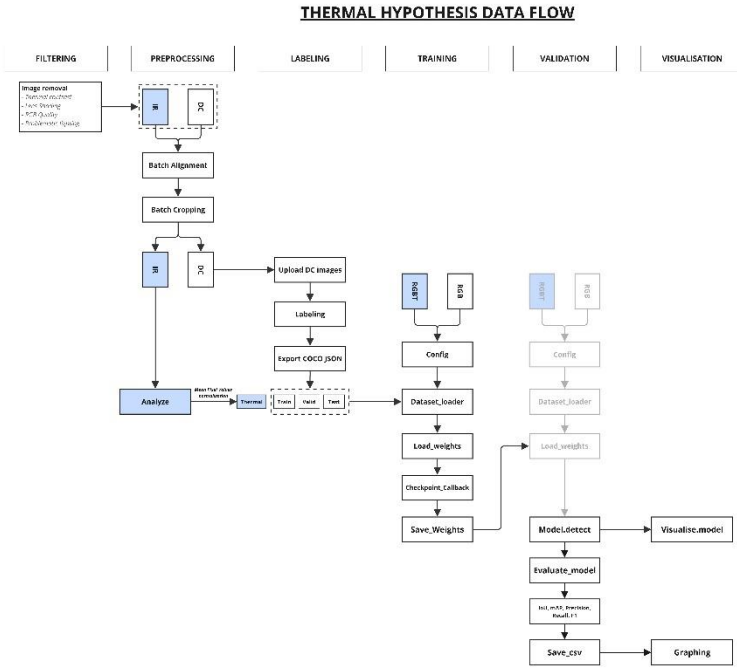
		Actual Class	
		Positive (**P**)	Negative (**N**)
Predicted Class	Positive (**P**)	True Positive (**TP**)	False Positive (**FP**)
	Negative (**N**)	False Negative (**FN**)	True Negative (**TN**)

The confusion matrix as shown in **table 7** quantifies how often one class is mistaken for another, making it particularly useful in experiments designed to reduce misclassification.

3.5. Model configuration and Data pipeline

H1 THERMAL | MODEL A RGBT vs Model B RGB

This section outlines the data processing and model configuration strategies used across all hypotheses. The process, from filtering to evaluation, is illustrated in the workflow diagrams (The Figure below for thermal hypothesis,).



DATA FILTERING AND PREPROCESSING

All datasets were initially filtered based on quality criteria to ensure consistency and reduce noise. Problematic samples were removed if they showed: Poor thermal contrast, Lens shading artifacts, Low RGB image quality, Inconsistent lighting (e.g., strong glare or shadows).

The remaining image pairs were then passed through an alignment and cropping pipeline. RGB and thermal (IR) images were aligned using one of two methods:

- Manual batch alignment using a fixed transformation with calibrated values (ScaleX = 0.880, ScaleY = 0.880, ShiftX = 93, ShiftY = 47, Rotation = -0.40°) for the T4xx series.

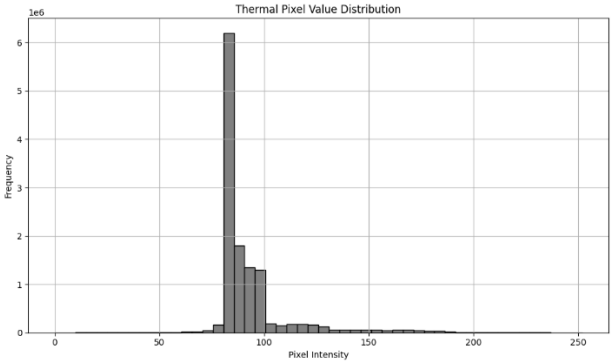
FLIR Thermal Studio export with fixed thermal min-max settings for the T96xx series. Following alignment, both channels were batch-cropped to remove edge padding. This preprocessing step ensured consistent geometry and scale for model training.

THERMAL PROCESSING AND FUSION INPUT

Thermal images were converted to grayscale and normalized using fixed pixel intensity bounds set in FLIR Thermal Studio. These normalized IR frames were then stacked as a fourth input channel for the RGBT model. The channel-wise MEAN_PIXEL values were calculated empirically and defined explicitly in the model configuration:

MEAN_PIXEL = np.array([123.7, 116.8, 103.9, 109.0])

This ensured that the thermal data was on the same scale as RGB input. This normalization was essential due to the limited range and resolution of thermal images.



LABELING AND DATASET STRUCTURE

Labeling was conducted using RGB images in COCO format. Each annotated image had a corresponding thermal file, linked by filename. Labels were manually drawn using polygon masks, and datasets were split into training, validation, and test sets. The same annotation file was used for both RGB and RGBT training by switching dataset loaders.

The dataset was labeled using the COCO (Common Objects in Context) JSON format, which supports polygon masks and multi-class annotations. This format enabled compatibility with the Mask R-CNN architecture and made it possible to run consistent training sessions across single-class (efflorescence only), multi-class (efflorescence + damage), and thermal fusion models. The same annotation file was reused for both RGB and RGBT pipelines, ensuring a 1:1 mapping of masks across modalities. Tools such as Roboflow were used to create and export the annotations.

MODEL ARCHITECTURE AND TRAINING STRATEGY

All experiments used a Mask R-CNN architecture with a ResNet-50 backbone. All models were initialized with COCO-pretrained weights using a transfer learning strategy. For RGB models, these weights were fully compatible. The training logic varied across models depending on the hypothesis:

Single-Class Models (Model A, B, C, H, J, K, L):
Trained to detect only efflorescence using either RGB or RGBT inputs.
Example:

```
NUM_CLASSES = 1 + 1  # background + efflorescence
```

Multi-Class Models (Model D, E, F):
Trained with two classes (efflorescence + damage type such as graffiti or lichens).
These models were critical to study misclassification risks and overlapping boundaries.

```
NUM_CLASSES = 1 + 2  # background + efflorescence + second class
```

Damage Co-Occurrence Models (Model H, I):
Explored spatial relationships between efflorescence and damaged bricks. The dataset was labeled with both damage and efflorescence masks and exported as a two-class setup.

Augmentation-Based Experiments (Model J, K, L):
Focused on robustness under varying lighting, rotation, and scale using the efflorescence-only dataset.

RGB-THERMAL (RGBT) MODEL CONFIGURATION

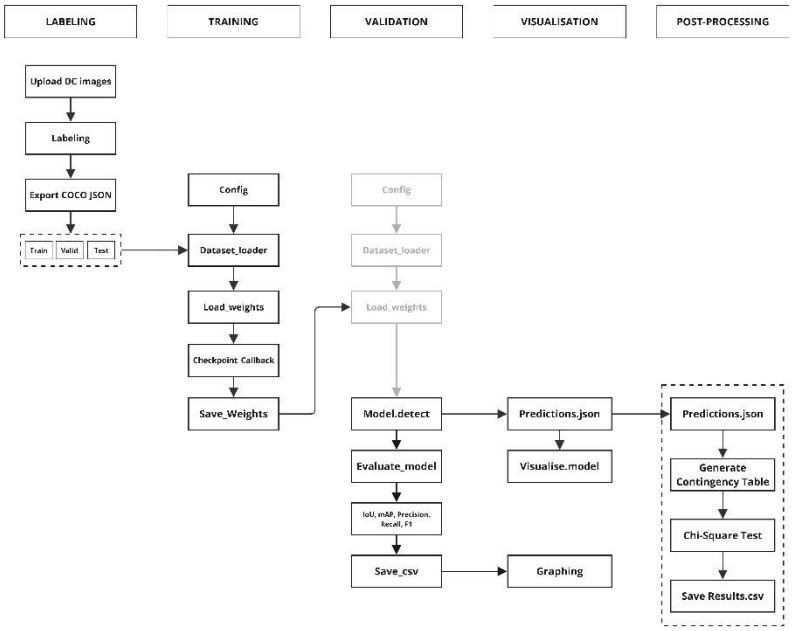
For the RGBT model (Model A), the input was extended to 4 channels. Since ResNet-50 does not natively accept a 4-channel input, the first convolutional layer (conv1) was skipped during training. This allowed the model to initialize thermal filters from scratch while reusing pre-trained weights for RGB features. During inference, this workaround was not needed as the full 4-channel tensor was accepted normally.

```
IMAGE_CHANNEL_COUNT = 4
BACKBONE = "resnet50"
IMAGE_MIN_DIM = 448
IMAGE_MAX_DIM = 448
USE_MINI_MASK = False
```

Weights were trained using callbacks with checkpointing. The training strategy included early stopping based on validation loss, and all losses (mask, class, and bbox) were logged per epoch.

POST-PROCESSING: DAMAGE CO-OCCURRENCE ANALYSIS

DAMAGE CO-OCCURANCE HYPOTHESIS DATA FLOW



In the co-occurrence hypothesis (H3), model predictions were exported as predictions.json files in COCO format. These were analyzed using a custom post-processing pipeline. The core goal was to detect spatial relationships, i.e., whether damage occurred on the same or neighboring bricks as efflorescence. A bounding-box-based approach was used, where detections were grouped into bricks, and neighbor relationships were computed using bounding box proximity and area overlap and eventually bounding box centroids.

For consistency in statistical testing, ground truth masks were used instead of predictions in the final analysis due to model instability in handling overlapping masks or multiple class assignments per brick. This provided a clearer basis to validate hypotheses about spatial co-occurrence patterns between damage and moisture-related efflorescence.

4. Experimental Results

This chapter presents the results of the model evaluations and hypothesis-driven experiments designed to assess efflorescence detection performance on masonry surfaces.

The analysis begins with a benchmark of two deep learning models (Mask R-CNN and YOLOv8) to establish a performance baseline. These models are trained on the same annotated dataset and compared using core training metrics such as bounding box loss and classification loss. Their performance is then evaluated using standard detection metrics: mean Average Precision (mAP@0.5), precision, recall, inference speed. Visual examples accompany the metric analysis to interpret qualitative differences in model behaviour.

Following this benchmark, **one model is selected** based on both quantitative performance and qualitative consistency. This model is then used throughout the remainder of this chapter to investigate four targeted hypotheses derived from literature and initial testing:

- **H1**: Incorporating thermal data as a fourth input channel improves the detection precision and segmentation accuracy of efflorescence, in moisture-related contexts, compared to RGB-only input.
- **H2**: Efflorescence will be visually misclassified more frequently when other similar looking surface changes (e.g., graffiti, lichens, encrustations) are present in the dataset.
- **H3**: The presence of contextual surface damage (e.g., powdering, scaling) increases the likelihood of efflorescence co-occurring in the same area.
- **H4:** Variations in image quality, angle, and distance negatively affect model performance in detecting efflorescence.

Each hypothesis is evaluated in a dedicated subsection using a combination of:

- **per-class detection metrics** (e.g., mAP per damage type),
- **confusion analysis** (to reveal misclassifications),
- and **visual inspection** of representative results.

This chapter thus aims to provide both a comprehensive comparison of candidate models and an in-depth understanding of the conditions under which efflorescence is reliably detected—or misinterpreted—by the selected model.

4.1. Global Performance Overview

This section presents the benchmark performance of two selected models (Mask R-CNN and YOLOv8) trained to detect efflorescence on masonry surfaces. These models were chosen based on findings from the literature and offer contrasting approaches: Mask R-CNN specializes in instance segmentation, while YOLOv8 is optimized for fast bounding box detection. The benchmark serves to determine which model is more suitable for further hypothesis-driven experimentation.

In supervised object detection and instance segmentation tasks, models are trained to minimize a set of loss functions that guide how accurately they localize, classify, and segment objects in an image. The two most relevant loss types are analyzed, **Bounding Box Loss** measures how accurately the model predicts the location of objects. It compares the predicted bounding boxes with the ground truth using metrics like L1 loss or smooth L1 loss. A lower bounding box loss indicates better spatial localization of objects. Classification Loss evaluates how well the model classifies the detected objects into the correct categories. It is typically computed using categorical cross-entropy. Lower classification loss means the model is assigning more correct labels to the detected objects.

In Mask R-CNN, these losses are computed separately for each region proposal and aggregated across the mini-batch. For YOLOv8, which uses a unified architecture, these components are part of an end-to-end optimization process that simultaneously considers box coordinates, class predictions, and objectness scores. Although mask loss is an important component of the Mask R-CNN architecture it is not included in this comparison. YOLOv8 does not perform instance segmentation in this study setup and only produces bounding boxes and class labels.

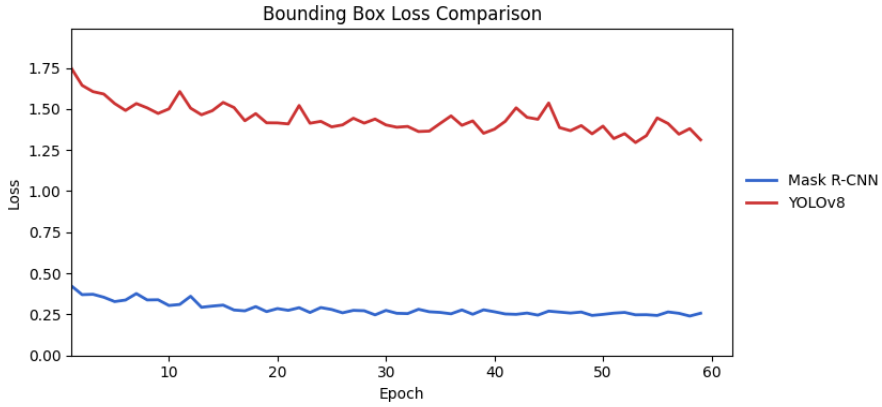


Figure 30 Bounding box regression loss for YOLOv8 (box_loss) and Mask R-CNN (val_mrcnn_bbox_loss). Lower values indicate more accurate localization of objects.

Across training epochs, Mask R-CNN consistently demonstrated significantly lower bounding box and classification losses compared to YOLOv8 as shown in figure 30 and 31. This indicates that its region proposal and classification heads are more stable and effective during learning. YOLOv8's losses, particularly in early epochs, are much higher but show a downward trend, suggesting it requires more time to stabilize. Nevertheless, even at later epochs, YOLOv8 fails to reach the low loss values exhibited by Mask R-CNN, which could be attributed to its anchor-free design and single-shot detection approach that emphasizes speed over precision.

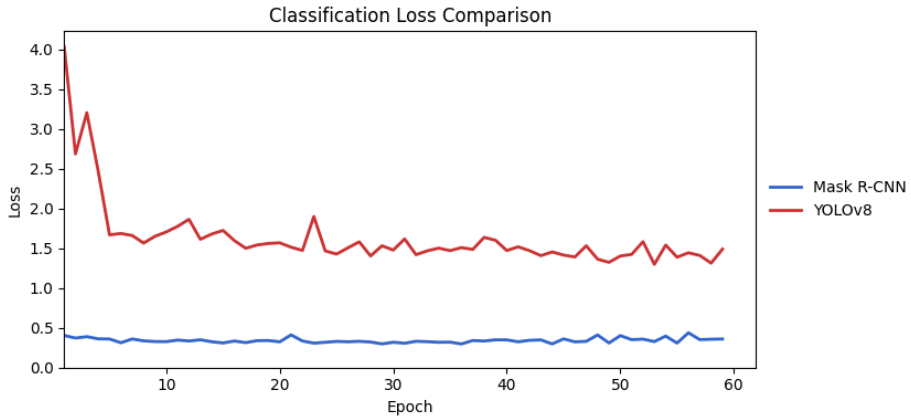


Figure 31 Classification loss over training epochs for both models. YOLOv8's class_loss is compared against Mask R-CNN's val_mrcnn_class_loss, showing how well each model learns to distinguish between classes.

Despite the lower losses of Mask R-CNN, the mAP@0.5 scores between the two models converge over time, with Mask R-CNN showing a slightly higher peak. Interestingly, YOLOv8 maintains consistently higher recall and slightly better precision, implying a stronger ability to detect more true positives overall while keeping false positives relatively low. This supports the idea that YOLOv8 is more aggressive in detection, while Mask R-CNN is more conservative and precise per instance.

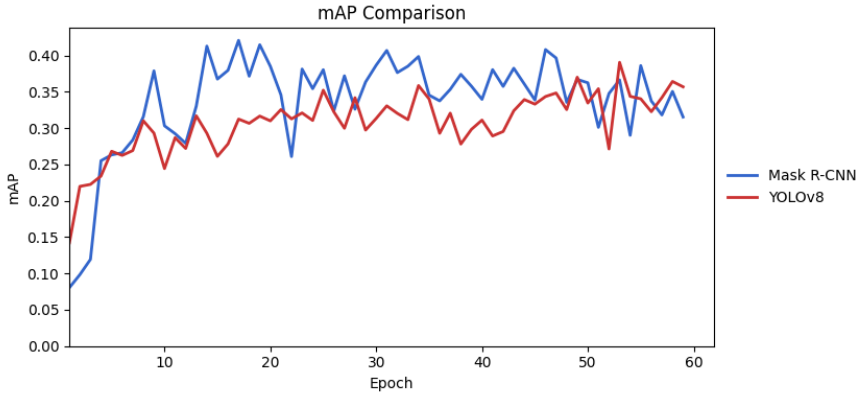


Figure 32 Comparison of mean Average Precision over 60 training epochs for YOLOv8 and Mask R-CNN. The graph illustrates the evolution of detection accuracy across training iterations.

Despite being trained for 60 epochs, both Mask R-CNN and YOLOv8 models demonstrate a relatively modest mAP score of 0.35, which is not considered optimal for high-confidence detection tasks. Interestingly, the similarity in performance between the two fundamentally different architectures suggests that the limitation may not lie in the model design, but rather in the dataset quality or structure itself. This reinforces the need to critically review and enhance the dataset by improving annotation consistency, increasing sample diversity, or balancing the class distribution, to enable the models to learn more effectively and achieve higher detection accuracy.

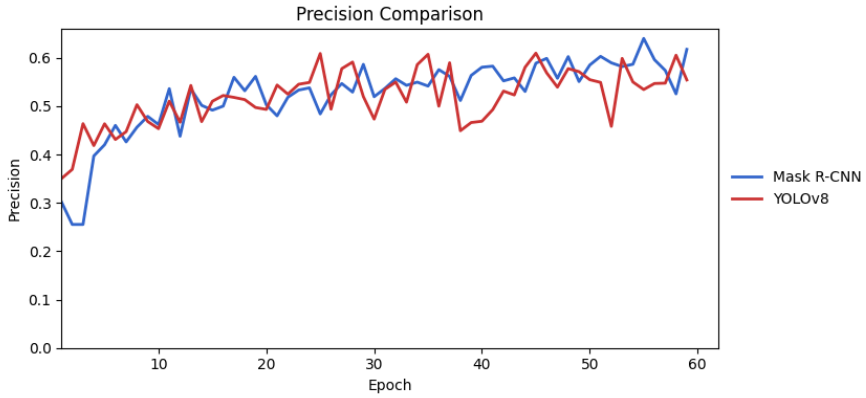


Figure 33 Precision values plotted over 60 epochs for both YOLOv8 and Mask R-CNN. Precision reflects the proportion of correct positive detections among all detections.

In addition to mAP, precision was evaluated as a complementary metric to capture how reliably the models distinguish true efflorescence from false positives. The precision score of around 0.60 suggests that the models are indeed on the right track, likely due to the detailed annotations, especially the marking of individual bricks exhibiting efflorescence. However, when comparing this performance to other studies, it becomes clear that there is room to enhance the model's output. For example, *Haixf et al. (2021)* achieved an mAP of 82.1% and a precision of 91.2% using Mask R-CNN on a more controlled dataset focused on efflorescence and related damage types. Similarly, *Kim et al. (2023)* reported an mAP of 0.884 and precision of 89.4% using YOLOv5, emphasizing damage classification including efflorescence.

These significantly higher performance metrics suggest that datasets with lower variability or domain-specific constraints tend to yield better outcomes. Therefore, the lower mAP in our study is not necessarily a failure of the architecture itself, but rather a reflection of the dataset's broad variability, environmental complexity, and possibly annotation inconsistencies. Future work may benefit from refining the dataset, either by increasing its size and consistency or by focusing on more homogeneous subsets of damage types to achieve performance gains comparable to prior research.

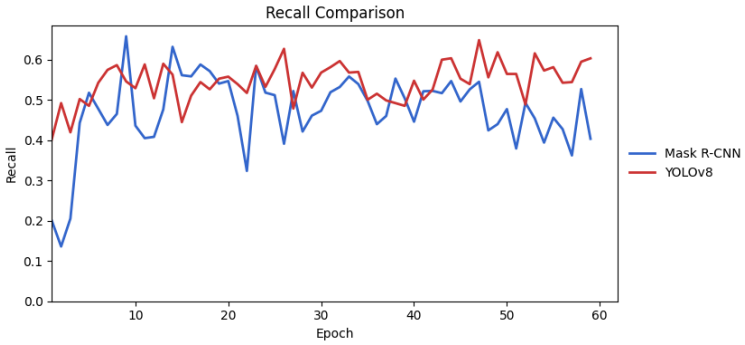


Figure 34 Recall performance comparison over training epochs. Recall indicates how well each model detects actual instances of the target class. Higher recall means fewer missed detections.

The sample image reveals an important limitation in the current detection process for example, some bricks with a whitish hue but no actual efflorescence are either falsely classified efflorescence, depending on the model. These visual ambiguities introduce a risk of misclassification, particularly for models like Mask R-CNN, which tend to be more liberal in generating segmented masks. This behaviour leads to the inclusion of visually similar bricks, even when they do not contain efflorescence. In contrast, YOLOv8 appears more conservative in its predictions, focusing on clearer,

more isolated efflorescence patches. This modesty in prediction likely results in greater precision, as observed in the metric comparison, albeit sometimes at the cost of missing finer instances.

Despite these nuances, both models demonstrate a broadly consistent pattern of detection, indicating that they share a general understanding of the damage signature. Nevertheless, the sample also underscores the need for continued refinement of both the dataset and model calibration, particularly in distinguishing efflorescence from benign discoloration or surface irregularities. Future enhancements should consider incorporating additional modalities or contextual cues to mitigate these classification challenges.

Figure 35 Side-by-side visual comparison of detection results on a test image. The left image shows the original input, the center shows Mask R-CNN results with instance masks, bounding boxes, class labels, and confidence scores, and the right image presents YOLO

The visual analysis reinforces the numeric trends. Mask R-CNN produces detailed segmentation masks with fine boundary alignment and higher confidence scores per instance. However, it often merges overlapping objects. YOLOv8, while lacking segmentation masks, delivers dense bounding boxes that more completely cover efflorescence areas, indicating better coverage but at the cost of precision and potential redundancy. YOLOv8's predictions are also stylistically more uniform, but less rich in spatial detail.

Table 6 Quantitative comparison between YOLOv8 and Mask R-CNN models. The metrics include precision, recall, mAP@0.5, average inference time per image (in milliseconds), and estimated frames per second (FPS). While YOLOv8 achieves slightly higher precision and

Model	Precision	Recall	mAP@0.5	Inference Time (ms)	Frames Per Second (FPS)
YOLOv8	0.60588	0.54806	0.35477	1565.93	0.64
Mask R-CNN	0.589112	0.46837	0.35652	1046.87	0.96

In terms of inference time, Mask R-CNN outperforms YOLOv8, achieving nearly 1 FPS versus YOLOv8's 0.64 FPS. Although both are not yet real-time, the faster execution of Mask R-CNN is notable given its instance segmentation capabilities. YOLOv8's higher latency is likely due to API overhead during remote inference, whereas Mask R-CNN runs fully locally.

4.1.1. Challenges and Limitations

During the benchmark phase, several technical, dataset-related, and model-specific challenges were encountered. Rather than being treated as general limitations of the entire study, these challenges provided critical insights that directly shaped the hypotheses tested in the following experiments.

A key constraint was the limited computational capacity available for training. The Mask R-CNN model, in particular, is memory-intensive and required substantial GPU resources. Due to restricted

VRAM on consumer-grade hardware, the batch size had to be reduced to 1, which significantly increased training time. Larger batch sizes caused memory overflows, particularly during high-resolution instance segmentation. To reduce risks of data loss, model weights were saved after each epoch. These constraints mainly affected the speed of experimentation rather than accuracy but highlight the practical barriers of applying advanced segmentation models outside high-performance environments.

The dataset also introduced several challenges that reduced generalization capacity. Small dataset size increased overfitting risks, while uncontrolled lighting conditions and shadows lowered prediction stability. Scale inconsistencies between annotated bricks and their real-world dimensions affected the model's perception of object proportions. Furthermore, partial or faint efflorescence annotations complicated ground truth definition and introduced label noise. These observations motivated Hypothesis 4, which tests how variations in image quality, angle, and distance affect model performance.

Figure 36 Comparison of detection results of non-efflorescence damages.

the original images (left column), Mask R-CNN predictions (middle), and YOLOv8 predictions (right) for five different classes:

graffiti, lichens growth, encrustation, exfoliation, and a second encrustation sample..

Another recurring difficulty was visual misclassification. Both Mask R-CNN and YOLOv8 struggled to distinguish efflorescence from visually similar phenomena such as graffiti, lichens, and encrustations

(Figure 36). Mask R-CNN produced dense over-segmentations, often including benign discolorations, while YOLOv8 was more conservative but still generated false positives. These issues motivated Hypothesis 2, which explicitly investigates the risk of misclassifying visually similar damage types.

In addition, efflorescence often appeared alongside other damage processes, such as powdering and scaling, complicating annotation consistency. This observation laid the groundwork for Hypothesis 3, which examines whether contextual surface damage influences the detection of efflorescence.

In summary, the benchmarking process revealed not only technical and dataset-related obstacles but also conceptual challenges in reliably distinguishing efflorescence. Rather than being treated solely as limitations, these observations directly informed the hypotheses that structure the subsequent experimental work.

Based on the benchmark evaluation, Mask R-CNN was selected as the primary model for the subsequent experimental phases. Although YOLOv8 demonstrated slightly higher recall and precision in some cases, Mask R-CNN's instance segmentation capability, its more stable loss convergence, and its suitability for pixel-level analysis made it the preferred choice for testing the hypotheses in the following chapters.

4.2. Hypothesis Evaluation

This chapter presents the experimental evaluation of the proposed hypotheses aimed at improving the detection of efflorescence in masonry surfaces using deep learning models. The primary goal is to assess how specific contextual and visual factors influence the model's performance. By systematically introducing variations in input conditions or dataset configurations, each hypothesis is examined independently while maintaining consistency in model architecture, evaluation metrics, and testing procedures. To ensure clarity and reproducibility, each hypothesis is presented in a consistent format, including an introduction, quantitative and qualitative results, and a brief discussion of findings.

This chapter aims to offer a comprehensive understanding of how each factor impacts efflorescence detection, guiding further improvements in the model and its practical applications.

4.2.1. H1: Infrared Thermal Imaging and Efflorescence Detection

This subsection addresses Hypothesis 1 (**H1**): *Incorporating thermal data as a fourth input channel improves the detection precision and segmentation accuracy of efflorescence, in moisture-related contexts, compared to RGB-only input.*

Efflorescence often appears in conjunction with moisture issues and manifests as white crystalline deposits on masonry surfaces. Moisture problems may originate from various sources, including rising damp, rain penetration, condensation, or leaks. Among these, two cases are especially relevant for this study because they produce clear and distinguishable patterns in both visual and thermal domains. Rising damp, caused by groundwater seeping upward through capillary action, often creates a colder region below the efflorescence due to evaporative cooling and persistent ground moisture. Leakage from above, for example through damaged gutters or roofing, typically results in colder zones above the efflorescence, reflecting recent water ingress from external sources.

These scenarios illustrate how thermal characteristics can diverge even when the visible patterns in RGB images appear similar. While both rising damp and leakage can lead to efflorescence, their distinct thermal signatures present an opportunity for models to interpret not only the visible appearance of damage but also its moisture-related thermal context.

To explore this, the model was trained on a set of approximately 150 RGBT images, each combining RGB channels with a thermal fourth channel. The goal was to assess whether this additional modality could help the model detect efflorescence more accurately, particularly in scenes where moisture-related thermal cues might reveal hidden or ambiguous damage patterns.

Unlike typical multi-class classification, the model was trained to identify efflorescence as a single class, leaving the thermal patterns for post-inference interpretation. The expectation was that the model would learn to associate efflorescence with its thermal environment, improving performance in cases where RGB-only models might struggle due to low contrast, surface discoloration, or subtle patterns.

The results below evaluate this hypothesis by presenting both quantitative metrics (loss functions, precision, recall, and mAP@0.5) and qualitative examples that compare RGB-only and RGBT predictions. These findings help determine whether thermal imaging meaningfully contributes to the detection and segmentation of moisture-driven efflorescence.

QUANTATIVE RESULTS

The loss curves from figure 37 show that the RGB model consistently achieves lower losses than the RGBT model across all components: total loss, bounding box loss, class loss, and mask loss. Particularly in the early epochs, the RGB model demonstrates a faster reduction in loss, suggesting a more efficient learning process. By epoch 60, both models converge, but the RGB model maintains a slight advantage in stability and final loss values.

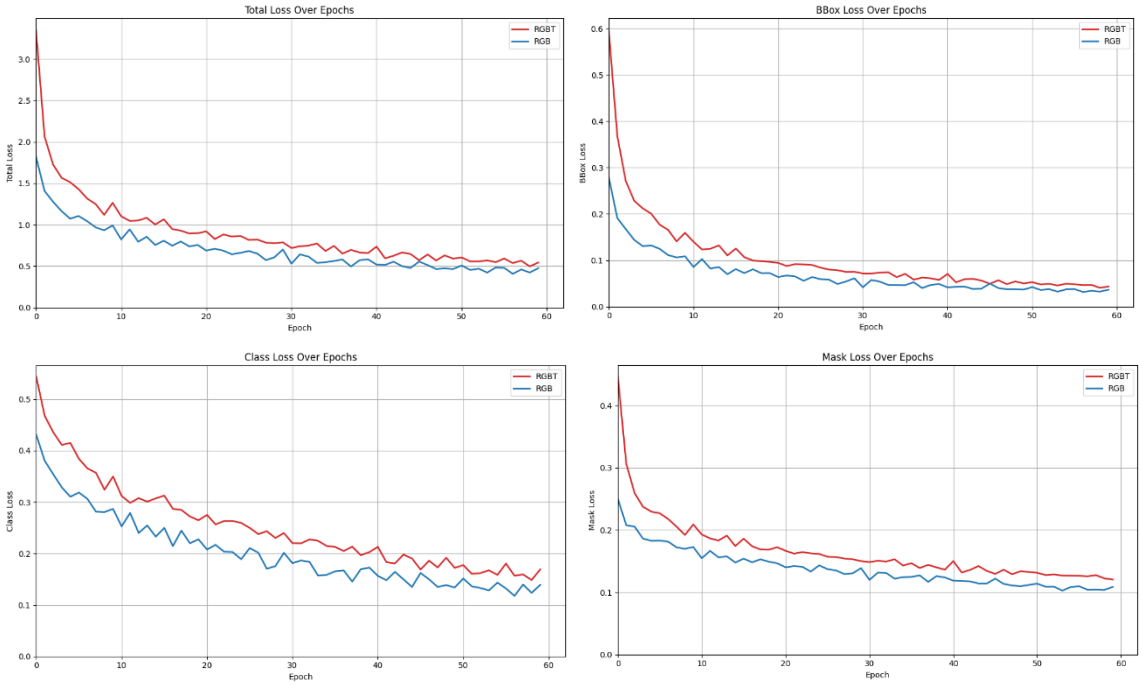


Figure 37 Loss functions per epoch over 60 training epochs comparing RGB-only (blue) and RGBT (red), (1) total loss, (2) box loss, (3) class loss, (4) mask loss.

The precision and recall plots illustrate distinct performance dynamics between the RGB and RGBT models. The RGB model shows consistently higher recall throughout training, indicating better detection coverage of true efflorescence instances. In contrast, the RGBT model demonstrates significantly higher precision, especially after epoch 10, suggesting it is more conservative but makes fewer false positive predictions. This trade-off between higher recall (RGB) and higher precision (RGBT) highlights their differing decision behaviours.

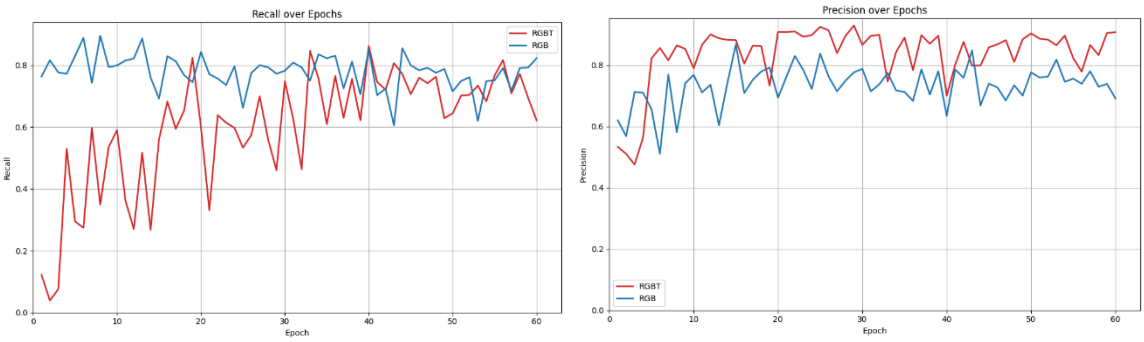


Figure 38 detection coverage comparison over 60 epochs for RGB-only (blue) and RGBT (red), (1) Recall, (2) Precision.

The mAP@0.5 plot demonstrates that the RGB model achieves faster and more stable convergence compared to the RGBT model. While RGBT gradually improves its performance over time, its trajectory is notably more erratic and less consistent, indicating instability during training. This

suggests that the fusion of thermal information introduces additional complexity, causing the RGBT model to struggle with convergence despite reaching competitive mAP levels in later epochs.

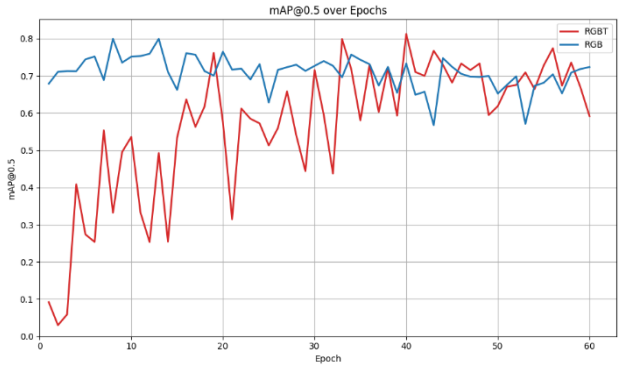


Figure 39 mAP@0.5 performance across 60 epochs for RGB-only (blue) and RGBT (red) models.

QUALITATIVE RESULTS

Model performance was evaluated by comparing RGB and RGBT predictions on 28 validation images, using metrics such as precision, recall, F1-score, mean average precision (mAP@0.5), and confidence. The RGBT model consistently achieved higher precision scores, particularly on complex images with high ground truth counts such as image 0328, where the RGB model achieved an F1 of 0.83 while RGBT improved this to 0.88.

Notably, RGBT predictions also maintained higher average confidence across the dataset (mean ≈ 0.96), suggesting greater certainty in predictions. In several cases, the RGB model suffered from elevated false negatives, indicating missed detections that RGBT was able to capture due to the added thermal information. This is evident in image 0065, where RGBT achieved a recall of 0.72 compared to RGB's 0.64. Overall, the addition of thermal data significantly reduced both under- and over-prediction errors, confirming its added value for robust efflorescence detection.

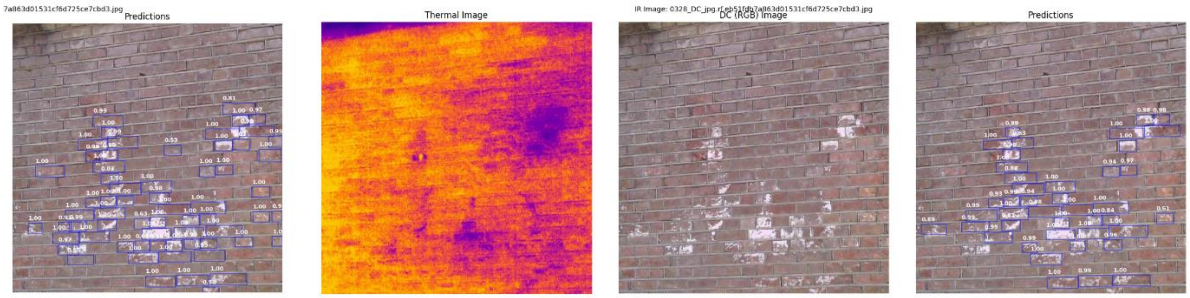


Figure 40: Efflorescence prediction results with confidence scores for Image 0328 (Epoch 60) (1) RGB Prediction. (2) Thermal Image. (3) Original Image: (4) RGBT Prediction

A comparative **table 7** shows the evaluation scores on figure 40 with image 328. In contrast, the RGBT model exhibited significantly fewer false positives (4), leading to a much higher precision (0.90). However, it missed more ground truth instances (11 false negatives), reflected in a lower recall of 0.78. Despite these differences, both models reached a similar F1-score (0.83 for RGB, 0.84 for RGBT), suggesting a trade-off between detection completeness and certainty.

Interestingly, the average prediction confidence is nearly identical (0.95 vs. 0.96), showing that both models are equally confident in their predictions. The results suggest that while the RGB model

captures more instances, the RGBT model is more conservative but precise — a potentially valuable trait in applications where false positives are more detrimental than occasional misses.

Table 7: Quantitative comparison of prediction performance for image 0328 (Epoch 60) between the RGB-only and RGBT models. Metrics include true positives (TP), false positives (FP), false negatives (FN), precision, recall, F1-score, mean average precision (mAP@0.5), and average prediction confidence.

Model	TP	FP	FN	Precision	Recall	F1	mAP	AvgConfidence
RGB	47	17	2	0.73	0.96	0.83	0.84	0.95
RGBT	38	4	11	0.9	0.78	0.84	0.76	0.96

Additionally for the 60th epoch the validation set is also analysed on the same metrics as can be seen in **table 8**. The RGB model exhibits a more balanced performance overall, with a recall of 0.86 and an F1-score of 0.77, compared to the RGBT model's recall of 0.65 and F1-score of 0.76. This suggests that RGB is more effective at capturing the majority of ground truth instances, despite a higher false positive rate (18 vs. 3).

On the other hand, the RGBT model shows a significantly higher precision of 0.94, indicating that its predictions are much more accurate when it does detect something — but at the cost of missing many true instances (22 FN vs. 11 FN). This leads to a lower mAP of 0.62 for RGBT, compared to 0.75 for RGB.

Interestingly, both models maintain high average confidence scores (0.97 for RGB, 0.94 for RGBT), meaning that each model is confident in its predictions regardless of its underlying tendency (recall-heavy for RGB, precision-heavy for RGBT).

Table 8 Quantitative average over the validation dataset comparison of prediction performance for (Epoch 60) between the RGB-only and RGBT models. Metrics include true positives (TP), false positives (FP), false negatives (FN), precision, recall, F1-score, mean average precision (mAP@0.5), and average prediction confidence.

Model	TP	FP	FN	Precision	Recall	F1	mAP	AvgConf
RGB	42	18	11	0.72	0.86	0.77	0.75	0.97
RGBT	31	3	22	0.94	0.65	0.76	0.62	0.94

DISCUSSION

The experimental results demonstrate that integrating thermal imagery as a fourth channel offers measurable advantages in the detection of efflorescence, particularly in terms of precision and confidence. The RGBT model consistently produced fewer false positives and maintained higher average confidence scores across both individual images and the full validation set. This confirms that thermal cues help the model make more certain and selective predictions, especially in complex or ambiguous scenes where RGB contrast alone may be insufficient.

However, the benefits of thermal data come at a cost. The RGBT model struggled with convergence, as indicated by more volatile loss and mAP curves. The added modality introduces complexity that challenges the training process, leading to inconsistent learning, especially in early epochs. Furthermore, the RGBT model exhibited a higher false negative rate, often missing subtle or faint efflorescence patterns that the RGB model was still able to capture. This indicates that while RGBT is more cautious, it may also overlook certain valid detections due to its conservatism.

From a performance trade-off perspective, the RGB model excels in recall and overall coverage, making it more suitable in contexts where missing any efflorescence is unacceptable — for instance, in preventive maintenance or early-stage diagnosis. In contrast, the RGBT model is more precise and confident, making it ideal in scenarios where false alarms carry greater operational costs, such as automated inspection systems or decision-support tools for restoration planning.

Lastly, the consistent performance across 60 epochs and the clear behavioral divergence between the models reinforce the importance of modality-aware model selection. The fusion of thermal and visual information holds promise, but its success depends on use-case priorities (e.g. recall vs. precision), model complexity, and data quality. Future work could explore hybrid strategies, for instance, using RGBT for initial detection and RGB for refinement, or vice versa.

Moreover, the observed instability and slower convergence of the RGBT model suggest that it may benefit substantially from extended training periods and larger datasets. The additional complexity introduced by the thermal modality likely requires more training iterations to reach full performance potential. It is expected that increasing the training duration to 110–150 epochs, combined with a broader and more diverse dataset, could improve convergence behavior and enable the RGBT model to better leverage thermal cues. Additional validation results supporting this are provided in Appendix I, under Model A (RGBT) and Model B (RGB).

4.2.2. H2: Misclassification Due to Similar Surface Features

The hypothesis examined in this section is that visually similar surface features, such as graffiti, lichens, and encrustation can lead to the misclassification of efflorescence. These features often share similar visual characteristics, such as white or light-colored patches, irregular textures, and surface deposits, which can confuse machine learning models. The goal of this experiment is to evaluate the model's ability to correctly distinguish efflorescence from these other surface features and to analyze the factors that contribute to false positives.

QUANTATIVE RESULTS

The training progress of the Mask R-CNN models for each damage class (Efflorescence, Graffiti, Lichens, Encrustation) is evaluated using four main loss components: total loss, bounding box loss, classification loss, and mask loss. Each of these contributes uniquely to the overall model performance and convergence behavior.

The total loss as described in **figure 41** shows a clear downward trend across all classes over 60 epochs, indicating consistent learning. Efflorescence demonstrates the most stable and lowest total loss towards the end, suggesting more reliable convergence compared to the other classes. Lichens and Encrustation show more fluctuation, which may reflect the higher complexity or variability in their visual appearance and segmentation.

Classification loss follows a generally declining pattern for all models. Model regarding Graffiti and Encrustation maintain the lowest class loss after epoch 20, suggesting that these models are confident in distinguishing their respective class from the background or other classes. The Efflorescence and Lichen models have slightly higher classification loss, likely due to visual overlap with other damages or subtle appearance. The mask loss, crucial for instance segmentation, decreases across all models. The Efflorescence model consistently performs best, with the lowest mask loss values. Lichen and Encrustation models exhibit more variability, again likely influenced by the fragmented and irregular mask structures typical of these damage types.

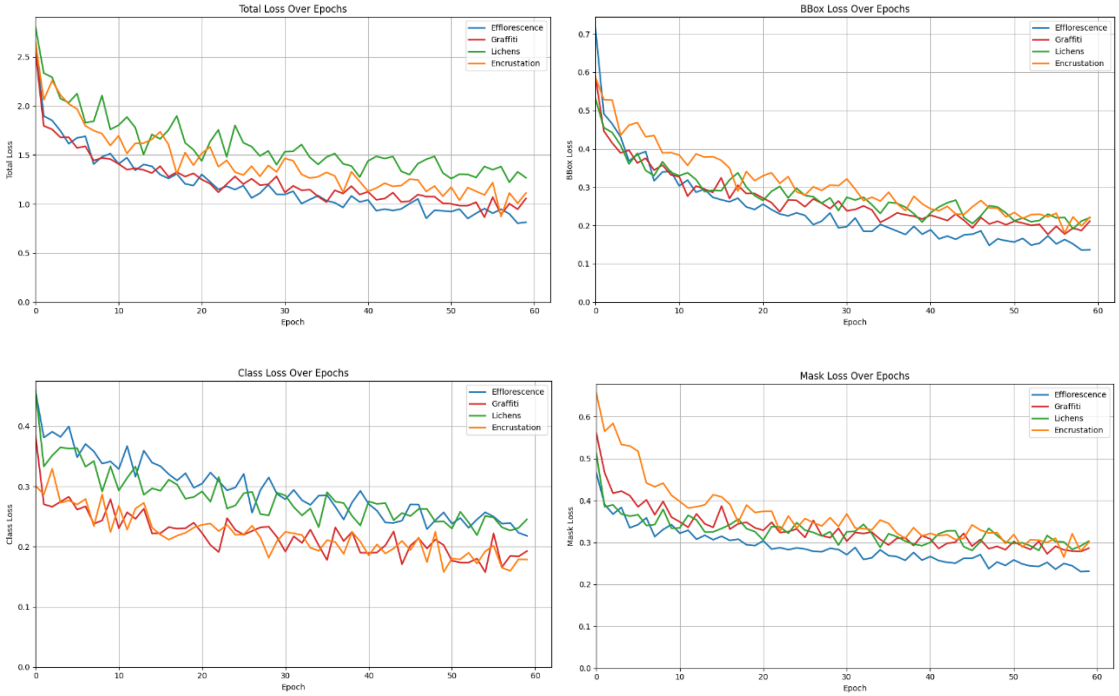


Figure 41 Loss functions per epoch over 60 training epochs comparing class specific Efflorescence (blue), Graffiti (red), Lichens (green) and Encrustation (orange), (1) total loss, (2) box loss, (3) class loss, (4) mask loss.

To evaluate the model's ability to differentiate between efflorescence and similar surface features, performance metrics were calculated, including mean Average Precision (mAP), precision, recall, and F1-score. In addition, a confusion matrix was generated to visualize prediction accuracy and error rates across different classes.

In reference to **figure 42,** Graffiti detection model consistently reached high precision and recall early in training, reflecting ease of detection due to distinctive visual features. Encrustation detection model improved over time and maintained reliable detection performance, though occasional drops in precision highlight some false positive fluctuations. Lichens detection model demonstrated high precision but somewhat lower and unstable recall. This suggests that when lichens were detected, predictions were often correct, but the model missed several actual instances (false negatives). Efflorescence detection model had the lowest and most fluctuating precision and recall. This is likely due to its variable appearance and overlap with other damage types, making it harder to detect consistently.

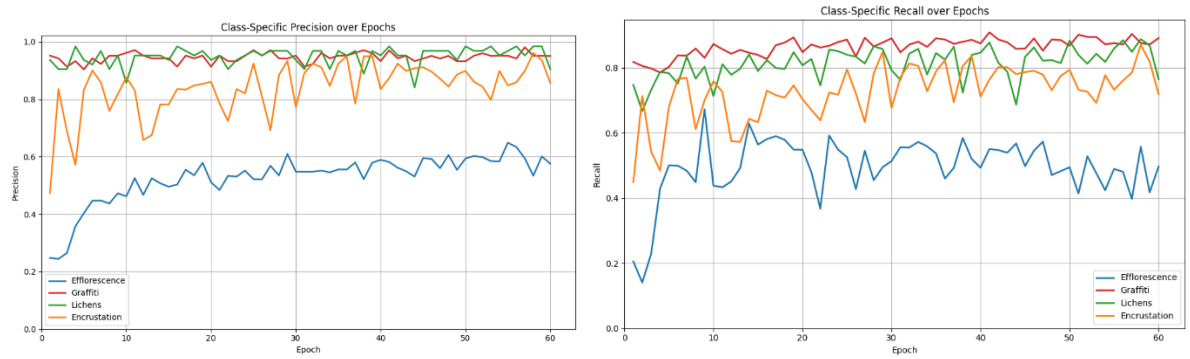


Figure 42 Precision and Recall5 performance across 60 epochs for Efflorescence (blue), Graffiti (red), Lichens (green) and Encrustation (orange)

The mean Average Precision at IoU threshold 0.5 (mAP@0.5) provides a robust measure of both classification and localization accuracy. As shown in **Figure 43**, detection quality increased for all damage types over training: Graffiti detection model achieved the highest and most stable mAP values, converging around 0.6, indicating consistent detections. Lichens and encrustation detection models showed gradual improvement, stabilizing between 0.45 and 0.5, though the fluctuations suggest some sensitivity to image variability or overlapping textures. Efflorescence detection model improved more slowly and plateaued below 0.45, reinforcing earlier findings that its diffuse and inconsistent appearance makes accurate detection more challenging.

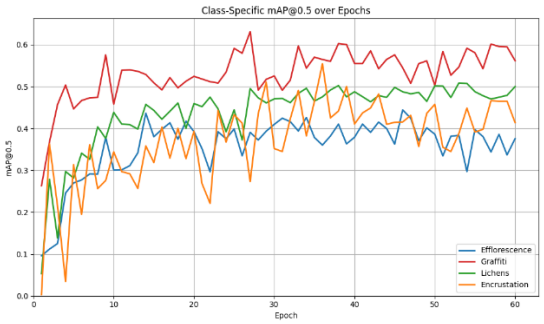


Figure 43 mAP@0.5 performance across 60 epochs for Efflorescence (blue), Graffiti (red), Lichens (green) and Encrustation (orange)

Overall, mAP@0.5 highlights how well each model balances precision and recall with spatial accuracy. While graffiti was detected most reliably, efflorescence remains the most difficult class due to ambiguous features and class confusion with other types of surface deterioration.

QUALITATIVE RESULTS

To evaluate the performance of models in distinguishing between visually similar damage classes such as efflorescence, graffiti, lichens, and encrustation, a simplified per-class, per-image confusion analysis was applied. Rather than relying on instance-level IoU-based matching, which is commonly used in object detection tasks, this approach counts true positives (TP), false positives (FP), and false negatives (FN) based on the presence or absence of predicted class labels in an image. This method was chosen due to several limitations associated with IoU matching in this specific context.

Damage classes like graffiti and lichens often appear with vague or irregular boundaries, leading to situations where one ground truth instance is predicted as multiple smaller segments, or a single large predicted mask spans several smaller ground truth regions. These discrepancies cause instance-level IoU values to fall below typical matching thresholds, even when the prediction is visually accurate. This results in valid detections being classified as false negatives. Conversely, if multiple predictions are matched to one ground truth object, the total number of true positives becomes inflated, distorting the results.

The simplified class-level method avoids these pitfalls by determining whether the correct class was detected in an image, regardless of the number or size of predicted masks. It provides a clearer representation of whether the model recognized the correct type of damage, which is essential when comparing performance across models trained to differentiate between similar-looking classes.

True negatives (TN) were intentionally excluded from the evaluation. Since each image contains only one damage class by design, counting TNs, such as not detecting graffiti in an image labeled as efflorescence, would introduce a large number of trivial "correct" non-detections. Including these

would skew the analysis and diminish the relevance of the actual classification performance between closely related classes.

GRAFFITI

The graffiti prediction demonstrates an oversegmentation behavior: the model generates multiple smaller masks within the same graffiti instance. Although visually correct—since they cover valid parts of the graffiti—these detections may not sufficiently overlap with the ground truth mask to surpass the IoU threshold (typically 0.5). As a result, they are counted as false positives rather than true positives, even when semantically accurate. This phenomenon aligns with observations in the precision-recall graphs, where graffiti showed high precision and recall, but occasional overcounting may inflate the FP count. Additionally, it relates to the Mask Loss plot, where graffiti maintains relatively low loss but still fluctuates, potentially due to these fragmented predictions.

In contrast, the efflorescence example shows large, confident masks with good coverage. The predictions are more consistent in shape and scale, corresponding to relatively stable and improving performance across all loss metrics. However, as seen in class-specific precision and recall plots, efflorescence maintains a modest precision and recall throughout training

Figure 44 Efflorescence and Graffiti predictions results with confidence scores (Epoch 60) (1) Original graffiti image. (2) Graffiti prediction. (3) Original Image: (4) Efflorescence prediction.

For the class Efflorescence, the confusion **table 9** shows a relatively balanced outcome with 31 true positives, 2 false positives, and 1 false negative. This suggests that the model has learned to detect efflorescence with reasonable accuracy. However, when comparing this to the loss curves, one aspect that stands out is the class loss, which remains higher for efflorescence throughout training compared

to other classes. This indicates that although the model can detect and localize efflorescence, it has difficulty assigning high confidence to its classification.

Table 9 Confusion matrix Efflorescence vs Graffiti

Class	TP	FP	FN
Efflorescence	31	2	1
Graffiti	67	1	2

ENCRUSTATION

The encrustation mask appears in figure 45 to underfit the actual damaged areas,while encrustation is broadly covered as a single continuous region it seems that the model generalizes to coarsely since the mask does not align with the damage type.

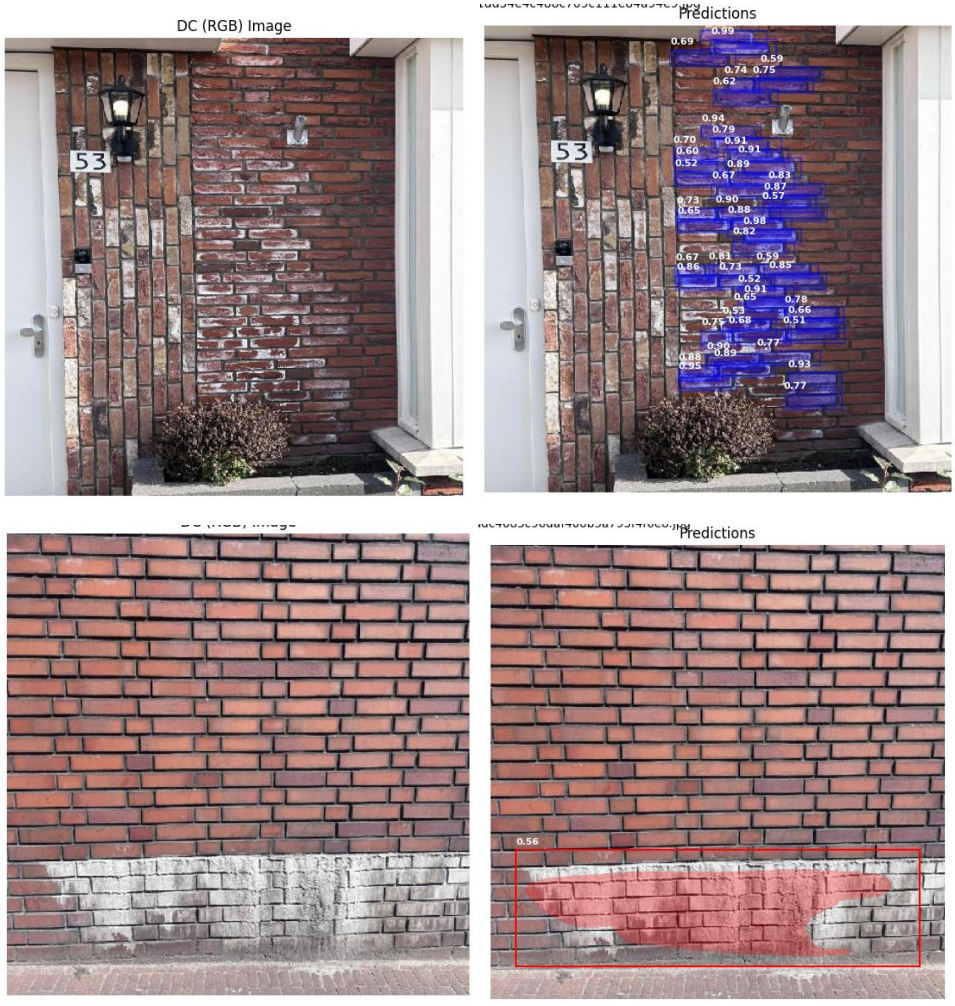


Figure 45 Efflorescence and encrustation predictions results with confidence scores (Epoch 60) (1) Original Encrustation image. (2) Encrustation prediction. (3) Original Image: (4) Efflorescence prediction.

Additionally, figure 45 reveals a false positive where the model mistakenly identifies a efflorescence area of the wall as encrustation. The consistency and repetition of such false positives point to either bias in the training data or limitations in the class-specific features learned. In cases like the image seen in figure 46, false positives frequently occur for encrustation. These are often triggered by surface textures, residual mortar, or lighting artifacts that visually resemble the rough surface patterns typical of encrustation but do not actually represent damage. This misclassification may be

amplified by the lower confidence thresholds used, as many false positives hover near the 0.5–0.6 score range. The model appears sensitive to textural patterns but fails to correctly distinguish them from true encrustation when context or edge definition is weak.

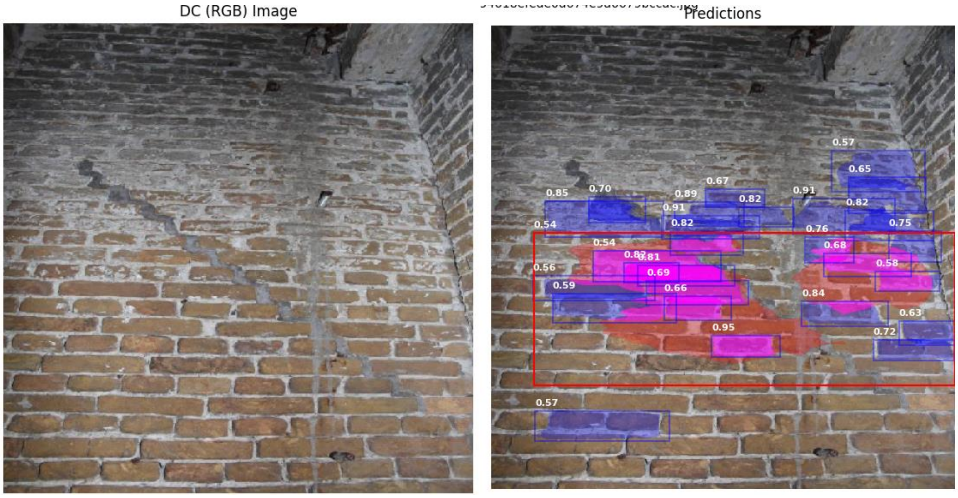


Figure 46 Efflorescence and encrustation predictions results with confidence scores (Epoch 60) (1) Original efflorescence image. (2) Efflorescence and (falsely) Encrustation prediction.

The confusion matrix presented in Table 10 indicates notable limitations in both precision and recall for the Encrustation class, consistent with the issues highlighted in the earlier visual analysis. While the model correctly identifies 36 true positives, it also produces 5 false positives and 4 false negatives. This pattern confirms that the encrustation class is currently over-predicting, it detects regions as encrustation that are not annotated as such, while also missing some genuine instances.

Table 10 Confusion matrix Efflorescence vs Encrustation at the 60th epoch

Class	TP	FP	FN
Efflorescence	33	1	0
Encrustation	36	5	4

LICHENS

Based on the results from the latest evaluation involving lichens and efflorescence, several observations can be made regarding model behavior and the influence of different loss components.

Figure 47 Efflorescence and Lichens predictions results with confidence scores (Epoch 60) (1) Original Lichens image. (2) Lichens prediction

From the confusion matrix in table 11, both efflorescence and lichens detection models demonstrate increased performance. The efflorescence detection model yields 33 true positives, 1 false positive, and 0 false negatives. The lichens detection model achieve a perfect score, 29 true positives, with zero false positives or negatives. This near-perfect result for both classes suggests effective localization, classification, and mask segmentation across both categories, but particularly for lichens.

When linking this outcome to loss functions, the results imply that both classification and bounding box regression losses were successfully minimized for the lichen class. This is supported by the dense and accurate bounding box predictions in the image of figure 47, each with high confidence scores and precise alignment with visible lichen patches. Moreover, the absence of false positives indicates that the model has learned good class separation in feature space, which reflects a well-optimized classification loss.

Table 11 Confusion matrix Efflorescence vs Lichens at the 60th epoch

Class	TP	FP	FN
Efflorescence	33	1	0
Lichen	29	0	0

DISCUSSION

The evaluation of the four classes, efflorescence, encrustation, graffiti, and lichens, reveals clear differences in performance, both in the visual predictions and in the training progression metrics. These differences reflect how effectively the model was able to learn and generalize the visual characteristics of each damage type.

The efflorescence detection model consistently shows strong performance. The model detects with high confidence, producing accurate masks that closely follow the extent of the damage. The results show minimal false positives and no false negatives in most cases, indicating that the class is visually distinct and well represented in the training set. During training, efflorescence steadily reaches stable loss and precision values, suggesting reliable learning and convergence.

The **encrustation detection model**, on the other hand, proves to be more challenging. Although the number of true positives is acceptable, the class exhibits significantly more false positives and false negatives. Visually, this is seen in underfitting masks that miss parts of the encrustation or simplify the structure too much. These issues suggest that the class is either visually inconsistent, shares features with other classes, or has weaker annotation clarity. *A key limitation is that the distinction between encrustation and efflorescence often depends on fine-grained surface detail, which is difficult to capture at lower image resolutions. More detailed, higher-resolution imagery would likely improve the model's ability to separate these visually similar classes.* In training, this challenge is reflected in fluctuating validation metrics and slower convergence.

The graffiti detection model performs relatively well and shows slightly better confidence scores compared to efflorescence in some cases. This may be due to the stronger color contrast graffiti typically has against masonry surfaces, making it easier for the model to distinguish. However, while graffiti benefits from these color differences, efflorescence remains limited in being falsely detected as graffiti. In most of the validation cases, efflorescence is either correctly classified or ignored entirely, rather than being misclassified, which confirms that the model maintains a strong internal representation of the class boundaries.

The lichens detection model emerge as the best-performing new class. The model detects them with high accuracy and confidence, without any false positives or false negatives. The predictions align

well with the actual growth patterns on the wall surfaces, and the segmentation masks are spatially accurate. This strong outcome is mirrored in training, where the lichen class reaches high performance quickly and maintains it across epochs. The visual distinctiveness of (yellow & white) lichens, especially their color (yellow) and texture, likely contributes to this performance.

In summary, while all classes are detected to some extent, efflorescence and lichens stand out as the most consistently learn ed. The graffiti detection model benefits from color-based distinction but still varies in mask quality, and encrustation requires more refinement due to visual ambiguity and segmentation challenges. These differences reinforce the importance of both annotation quality and visual distinctiveness in training effective damage detection models.

4.2.3. H3: Co-Occurrence of Damage Types

The third hypothesis investigates whether the presence of other types of surface damage can influence the model's ability to detect efflorescence. Efflorescence, as previously described, refers to white salt deposits appearing on masonry surfaces, often signaling moisture-related degradation. In contrast, damage in this study refers to various physical manifestations such as disintegration, loss of cohesion, layering, and other masonry deterioration types, as categorized in the MDCS and discussed comprehensively in Section 2.3.

To investigate potential relationships between efflorescence and nearby damage, a Mask R-CNN model was trained to detect both classes, efflorescence and damage, on annotated masonry imagery. This enables an automated analysis of spatial proximity and potential co-occurrence patterns across a diverse set of conditions.

To test whether there is a relationship between the presence of efflorescence and surrounding damage, the following hypotheses were formulated:

H_0 (Null Hypothesis):

The presence of efflorescence in a brick is independent of damage in its neighbouring bricks.

H_1 (Alternative Hypothesis):

Bricks near efflorescence are more likely to show signs of damage than bricks located further away.

While model predictions provide a useful estimation of damage and efflorescence presence, they may be affected by class imbalances, limited data, and segmentation uncertainty. To strengthen the reliability of this statistical relationship, the co-occurrence analysis is therefore based on the ground truth annotations rather than on predicted masks.

Each annotation (bounding box) is analyzed based on its centroid position, and co-occurrence is evaluated by checking whether a mask of one class overlaps with the centroid of another. This approach allows flexibility across varying image scales and avoids hardcoded pixel thresholds, which might fail in zoomed or low-resolution cases.

To account for varying spatial relationships, distances between centroids are normalized using the average annotation width per image. Based on this, several distance zones are defined:

- Zone 1: within 2× average width (same or neighboring brick)
- Zone 2: within 3× average width
- Zone 3: further than 4× average width

- Etc.

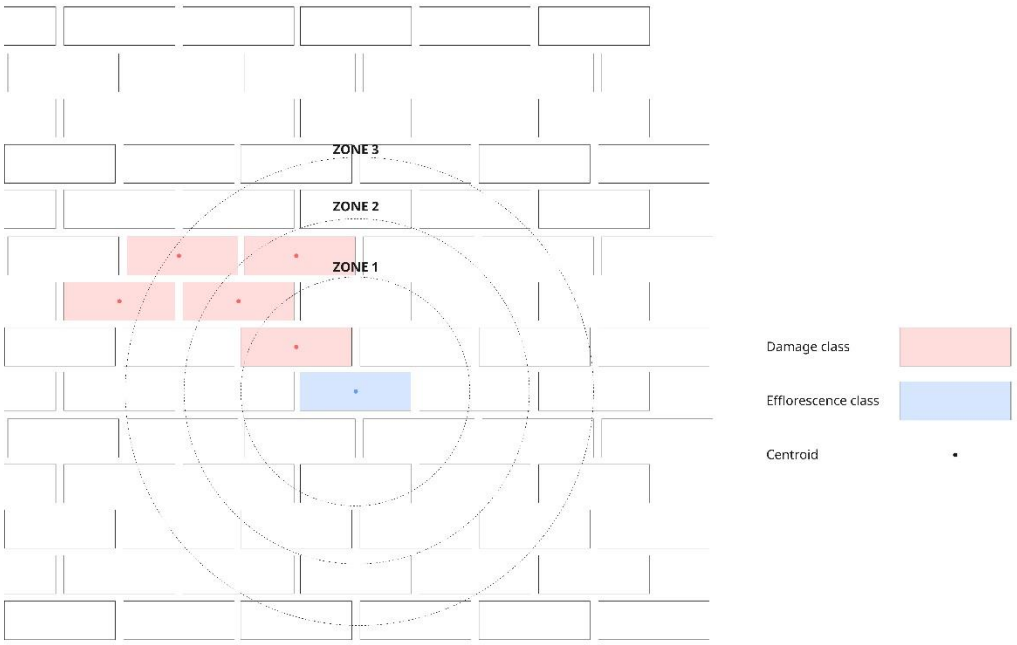


Figure 48 Illustration of the centroid–overlap method used to analyze spatial co-occurrence between efflorescence and damage annotations. Blue represents efflorescence and red represents damage. Distances between centroids are normalized by the average brick width, defining three zones: Co-occurrence is recorded when a damage centroid falls within the defined zones around efflorescence.

This stratified distance framework as shown in figure 48 allows us to investigate whether the probability of damage decreases as distance from efflorescence increases, thus supporting or rejecting the alternative hypothesis.

The analysis will involve the following: Constructing a contingency table of damage occurrences at various distances from efflorescence, Performing a Chi-Square test for independence to assess whether proximity to efflorescence is associated with a higher likelihood of damage, Optionally, visualizing damage probability as a function of distance to identify trends.

This approach aims to go beyond visual inspection, providing a statistically grounded insight into the co-location of salt-induced damage phenomena in masonry structures.

QUANTATIVE RESULTS

The Mask R-CNN model was trained for 60 epochs using the annotated dataset containing instances of efflorescence and damage. The training progress was monitored using standard loss functions: total loss, bounding box (BBox) loss, classification loss, and mask loss.

As shown in the training logs in figure 49, the total loss decreases steadily during the first 20–30 epochs, indicating that the model is learning meaningful patterns. After that point, the loss plateaus around a value slightly above 1.0, suggesting limited further improvement. The bounding box loss converges rapidly to below 0.1, indicating that the model can localize objects with reasonable accuracy. However, the classification loss stabilizes at a relatively high value (0.5), pointing to challenges in distinguishing between efflorescence and damage. This may stem from visual overlap, limited training examples, or imbalances in class frequency. The mask loss also shows initial improvement but converges around 0.15–0.2, which suggests that segmentation quality is moderate but could be improved with more data or refined annotations.

In summary, while the model is able to learn spatial and visual cues for detection, particularly for efflorescence, the performance on damage classes remains suboptimal. This justifies the decision to base the statistical co-occurrence analysis on the ground truth annotations rather than relying solely on model predictions.

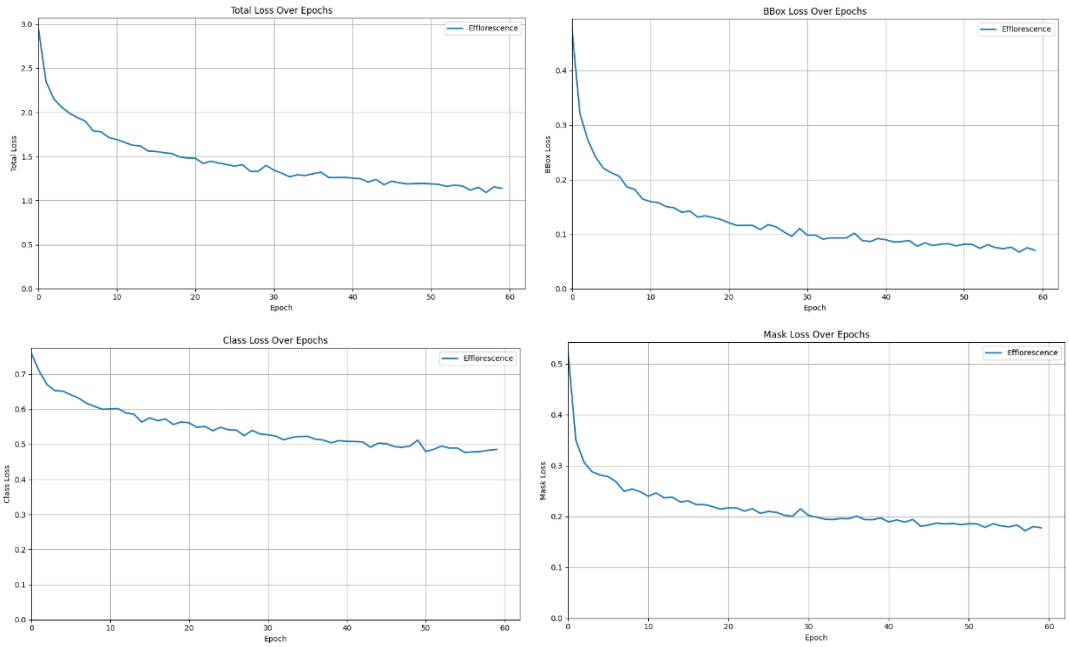


Figure 49 Loss functions per epoch over 60 training epochs for Efflorescence and damage class training), (1) total loss, (2) box loss, (3) class loss, (4) mask loss.

The evaluation results for the combined class of Efflorescence & Damage are shown in figure 50. While recall shows a generally increasing trend throughout the training process, reaching values around 0.40, it remains relatively low, suggesting that the model struggles to detect all true instances of this class. The fluctuation indicates instability in learning, likely due to class imbalance or the complexity of accurately detecting co-occurrence cases.

In contrast, precision initially starts high (above 0.7) but gradually declines and stabilizes between 0.5 and 0.6. This indicates that although many predictions are correct, the model becomes less selective over time, possibly generating more false positives as training continues. This inverse relationship between precision and recall is common in underrepresented or noisy classes.

These results emphasize the challenge of detecting both damage and efflorescence in the same instance, especially when the two classes overlap visually or contextually. The unstable pattern suggests that further dataset refinement or class balancing may be necessary to improve consistent detection performance.

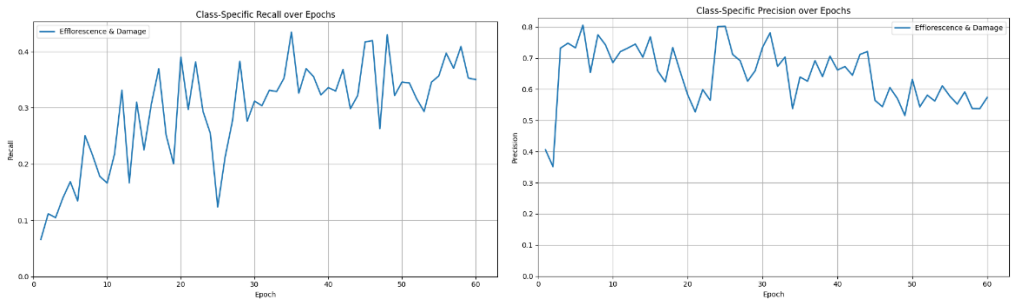


Figure 50 detection coverage over 60 epochs for Damage & efflorescence detection (1) Recall, (2) Precision.

Figure 51 presents the mean Average Precision (mAP@0.5) across training epochs for the classes Efflorescence and Damage. The damage class consistently outperforms the efflorescence class, reaching peak mAP values of around 0.45, while efflorescence stabilizes closer to 0.30–0.35. This performance gap suggests that the model is more confident and accurate in localizing and classifying damage compared to efflorescence.

Both classes show fluctuating performance across epochs, which is indicative of training instability — likely due to the relatively small dataset size, annotation density (some images have hundreds of annotations), and potential class imbalance. Notably, efflorescence shows a dip around epoch 25, after which it recovers but does not exceed its earlier peaks.

Overall, while damage detection appears more robust, efflorescence detection remains more challenging. This discrepancy justifies further post-processing and co-occurrence analysis using ground truth annotations rather than relying solely on predictions.

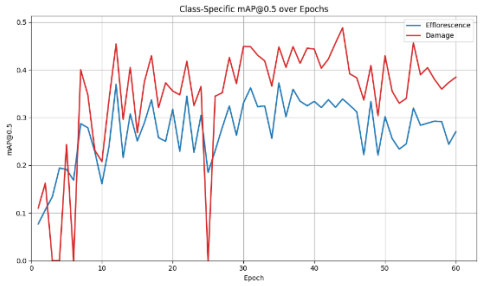


Figure 51 mAP@0.5 performance across 60 epochs for Efflorescence and Damage detection models

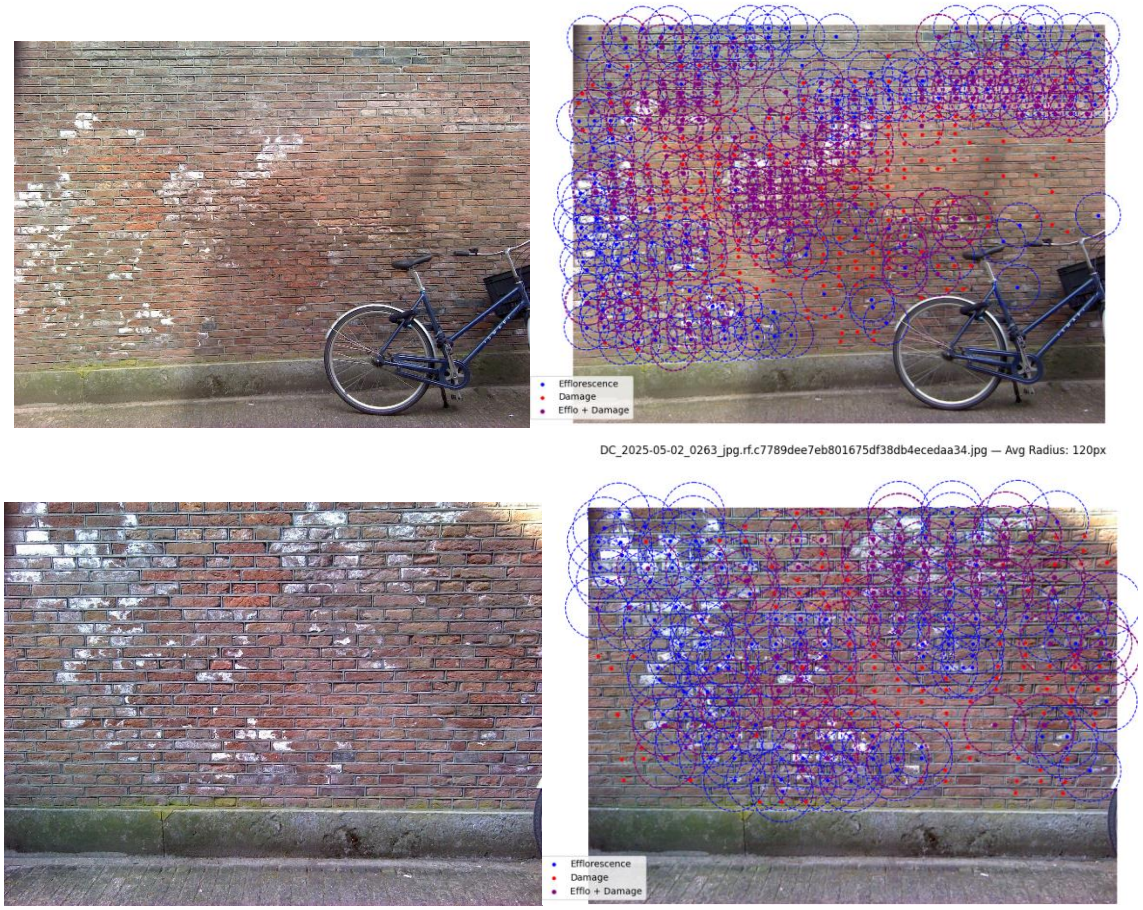


Figure 52 Radius determination in relationship with efflorescence (blue), Damage (red) and daamge & efflorescence in the same brick (purple) with radius

QUALITATIVE RESULTS

The comparison between the model's predictions (right) and the original RGB image (left) in figure 52 and 47 highlights key challenges in the detection of overlapping efflorescence and damage. While the model is capable of identifying individual regions with either efflorescence or damage (as indicated by blue and red bounding boxes), it appears to struggle with correctly classifying bricks that exhibit both phenomena simultaneously. This may reflect the complexity of visual overlap and the model's sensitivity to subtle patterns. Nevertheless, the overall shape and placement of predicted masks and bounding boxes remain consistent with visual observations and annotations, indicating a solid spatial understanding of both damage and efflorescence separately.

Figure 53 Efflorescence and damage predictions results with confidence scores (Epoch 60) (1) Origina image. (2) Efflorescence and damage prediction.

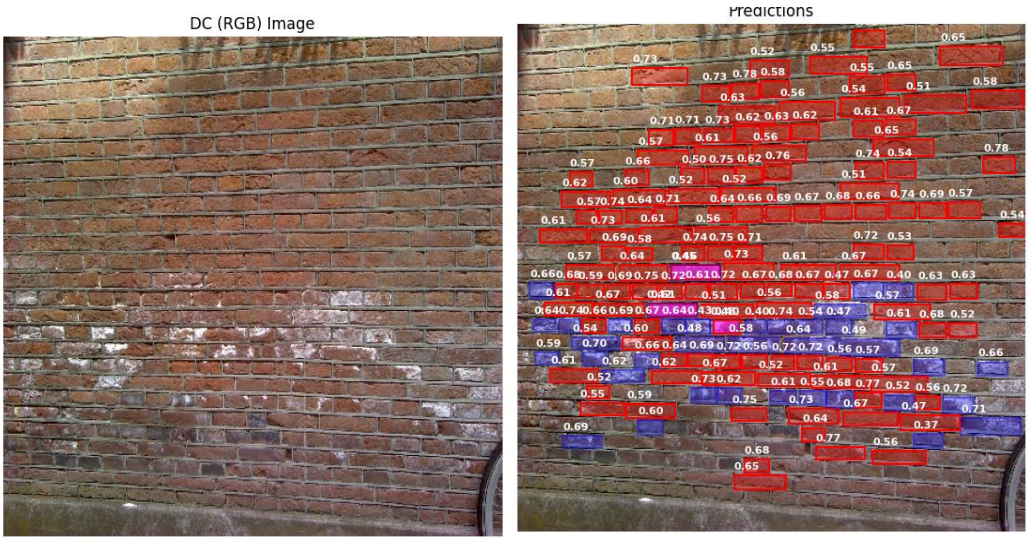


Figure 54 Efflorescence and damage predictions results with confidence scores (Epoch 60) (1) Origina image. (2) Efflorescence and damage prediction.

DISCUSSION

This study explored the spatial relationship between efflorescence and masonry damage (including disintegration, loss of cohesion, layering, etc., as introduced in Chapter 2.3 and described in the MDCs). The goal was to explore whether such co-occurrence analysis (using ground truth + statistical testing) can increase confidence and precision in detection frameworks. To investigate this, a chi-square test of independence was performed across varying spatial neighbourhood distances, using ground truth annotations instead of model predictions due to limitations in the detection performance.

Statistical Findings on Spatial Co-occurrence

The results from the chi-square test for different radius factors are summarized in table 12:

Table 12 Results of chi-square test on Efflorescence near damaged bricks and their relationship

Radius Factor	Radius (px)	Chi-square	p-value
1	114	47.13	< 0.00001
1.5	171	47.13	< 0.00001
2	228	47.13	< 0.00001
5	570	6.77	0.00927

These results show a statistically significant association between damage and the nearby presence of efflorescence for radius factors between 1.0 and 2.0. Even at a radius factor of 5.0, the relationship remains statistically significant, although the p-value increases substantially. This suggests that damage tends to co-occur with efflorescence within a localized spatial range. This supports the hypothesis (H_1) that bricks near efflorescence are more likely to show damage for this specific case. Additional data such as the contingency tables, Chi square per cut-off are added in APPENDIX I

While the Mask R-CNN model was trained to detect both efflorescence and damage, visual inspection of prediction results showed a key limitation: efflorescence was frequently not detected in bricks already marked as damaged, despite being visible in the ground truth annotations. The masks and bounding boxes for both classes were generally accurate when they occurred separately, but co-located cases (efflorescence overlapping with damage) were often missed.

This shortcoming motivated the use of manual (or Ground Truth) annotations instead of model predictions in the statistical analysis. The prediction results would have introduced significant bias, underrepresenting the true co-occurrence frequency and weakening the statistical results. This emphasizes the importance of high-quality annotation and dataset balance in training, especially when modeling overlapping or visually damage phenomena.

The findings confirm a significant spatial correlation between efflorescence and damage, especially in closer proximity. The limitations in detection performance further highlight the challenge of learning co-occurring features in deterioration patterns and suggest the need for better multi-label segmentation approaches or dataset augmentation strategies.

4.2.4. H4: Image Quality, Camera Angle, and Scale

The fourth hypothesis explores how variations in image quality, camera angle, scale, and lighting conditions affect the accuracy of efflorescence detection. Capturing high-quality images under consistent conditions is often challenging in field applications, leading to variations that can degrade model performance. In particular, low resolution, steep viewing angles, differing scales, and poor lighting can obscure key features necessary for accurate detection. The following hypothesis was formulated **H4:** Variations in image quality, angle, scale and distance negatively affect model performance in detecting efflorescence. This hypothesis aims to determine whether augmenting the training dataset to simulate these variations can improve the model's robustness and generalization

To test this hypothesis, datasets were systematically augmented to simulate different acquisition conditions. Each augmentation type was applied to the validation subset, while ground truth masks were transformed alongside the images to ensure perfect alignment. The augmentation types included:

- Camera angle variation (±5°, ±10°, ±15°).
- Scale variation (0.5×, 0.75×, 1.25×, 1.5×).
- Resolution variation (480px, 320px, 160px, rescaled to 640px).
- Lighting variation (±10%, ±20% in brightness and contrast).

The model was evaluated on each augmented dataset over 60 epochs, and results were compared against the baseline dataset.

QUANTATIVE RESULTS

Figure 55 shows the model's performance across simulated angular deviations. Mild angle changes (±5°) produce minimal performance drops compared to baseline. However, as angular deviation increases (±10°, ±15°), especially in the ±15° condition, mAP drops significantly and training becomes less accurate. Dashed lines (positive angle deviations) and solid lines (negative) show similar trends, suggesting angle direction has little effect. These results imply that the model is moderately robust to small viewpoint shifts but becomes less reliable under stronger tilt conditions above 15 degrees.

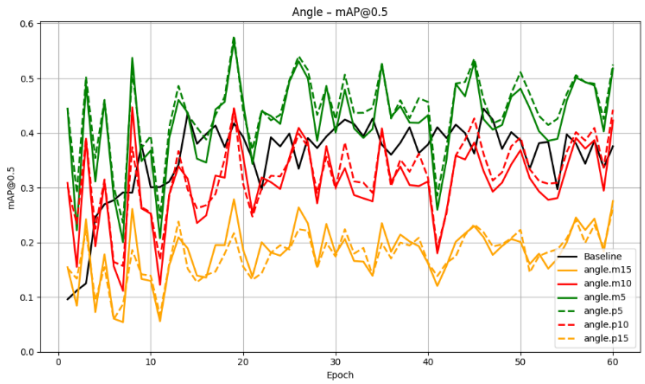


Figure 55: mAP@0.5 per epoch for varied camera angles (±5°, ±10°, ±15°) compared to the baseline.

Figure 56 compares detection performance across rescaled inputs. Downscaling (0.5× and 0.75×) leads to sharp declines in mAP, indicating that reduced image detail hampers model learning. Upscaling (1.25×, 1.5×) shows mixed results, 1.25× approaches baseline performance, but 0.5× becomes unstable, likely due to scaling artifacts and small mask predictions. These results confirm that extreme resizing can degrade model performance, either by blurring detail or introducing distortions.

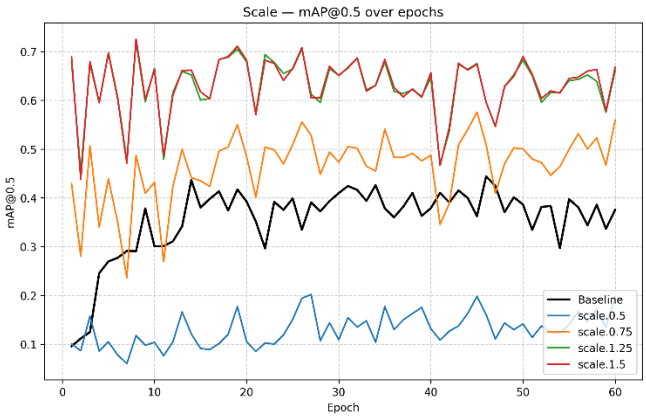


Figure 56 mAP@0.5 per epoch for images rescaled by 0.5×, 0.75×, 1.25×, and 1.5×, relative to baseline resolution

In Figure 57, 3 out of 4 lighting-altered variants show increased detection performance compared to the baseline. Especially in the extreme cases (-20%), the model fails to achieve stable learning, with mAP values remaining consistently low. Less intense lighting changes (-10%) perform slightly better but still underperform compared to baseline. Most interesting is the (+10%) lighting shift where is

significantly improved detection performance. These findings indicate that the model is sensitive to exposure changes and may benefit from more robust color normalization or lighting-invariant feature learning.

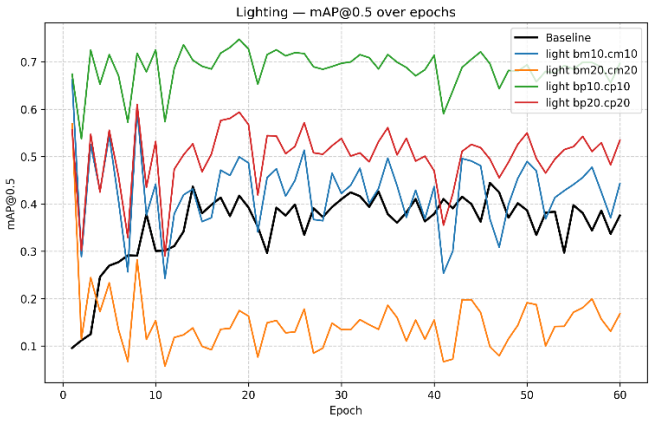


Figure 57 mAP@0.5 per epoch for images with simulated brightness/contrast shifts (±10%, ±20%), compared to baseline.

Figure 58 presents the mAP@0.5 across epochs for different input resolutions. The baseline model shows similar result as the 160 resolution-modified variant, especially during end and mid-training. The 480 and 320 resolution settings achieve relatively close performance, and both plateau approximately 20% above the baseline's peak. The lowest resolution (160) shows significantly reduced performance at the beginning, with unstable training and lower mAP. This suggests that resolution degradation particularly affects the model's ability to localize fine-grained efflorescence features.

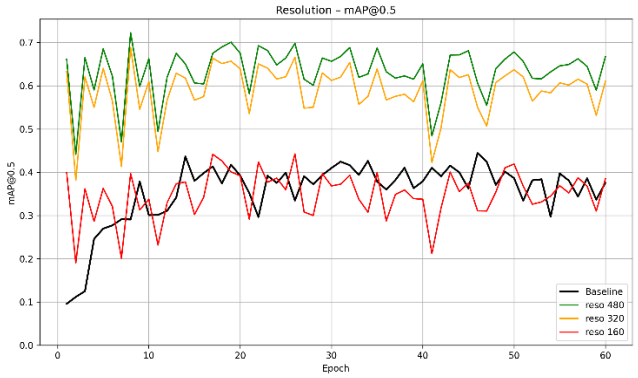


Figure 58 Mean Average Precision (mAP@0.5) per epoch across different image resolutions (160, 320, 480), compared to the baseline model.

QUALITATIVE RESULTS

In addition to quantitative evaluation, a qualitative assessment was conducted to examine how various augmentation parameters affect the model's detection behavior on a consistent visual input. A single validation image was selected as a baseline, and this image was then augmented using each of the different conditions tested during training: **viewpoint (angle) shifts**, **scale variations**, **lighting changes**, and **resolution reductions**. For each variant, predictions were generated using the model checkpoint from **epoch 60**, enabling a direct visual comparison of segmentation outputs across conditions.

This setup allows for side-by-side inspection of how prediction masks shift, degrade, or improve under each transformation, providing valuable insight into the spatial reliability and visual coherence of the model's segmentation capabilities.

Figure 59 base line image 650x640 px

ROTATION

As shown in Figure 60, the model demonstrates moderate robustness to changes in camera tilt. In the ±5° and ±10° rotation cases, most efflorescence bricks are still detected, and mask coverage remains consistent, although confidence scores start to drop slightly. However, at ±15°, several key issues emerge:

- Some detections shift slightly from their correct positions, especially near the image edges.
- A few smaller bricks are missed altogether.
- Confidence values drop below 0.80 in multiple areas, signaling increased uncertainty.

Despite these issues, the model does not show major asymmetry between counterclockwise tilts rotational direction seems to have negligible effect, which aligns with the symmetrical performance trends seen in the mAP curves. This suggests that moderate angular shifts are tolerable, but stronger tilt begins to interfere with spatial consistency and mask placement.

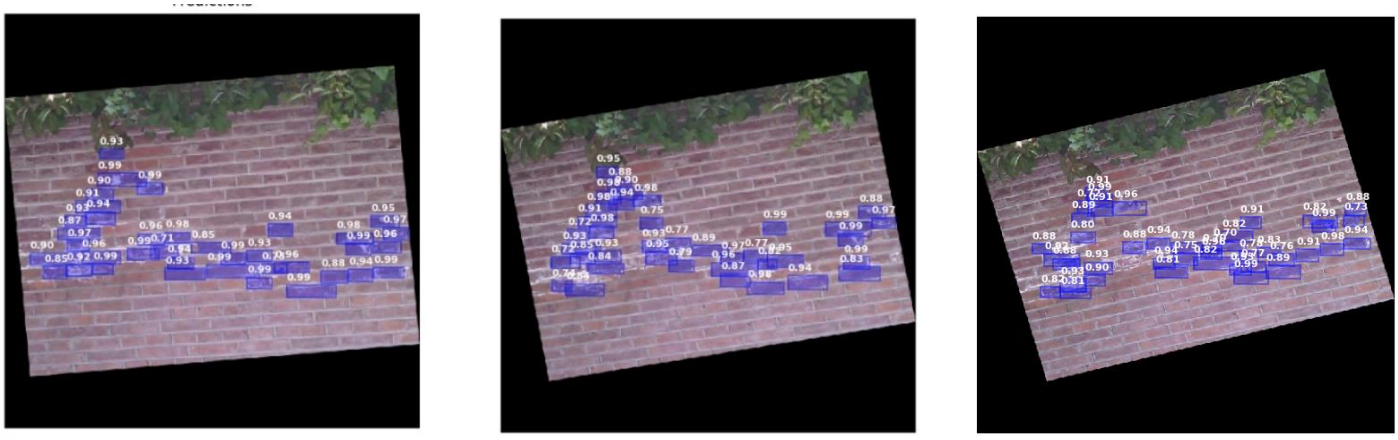


Figure 60 Predicted masks under input rotation (5°, 10°, 15°) compared to the baseline.

SCALING

Downscaling the input (0.75×, 0.5×) leads to clear losses in mask resolution. Efflorescence is under-segmented, and minor patches disappear entirely. Upscaling (1.25×, 1.5×) slightly enhances boundary detail but also introduces noise around edges, particularly at 1.25×. The base image 1.5× variant yield the most visually coherent results, highlighting a trade-off between scale enlargement and noise amplification.

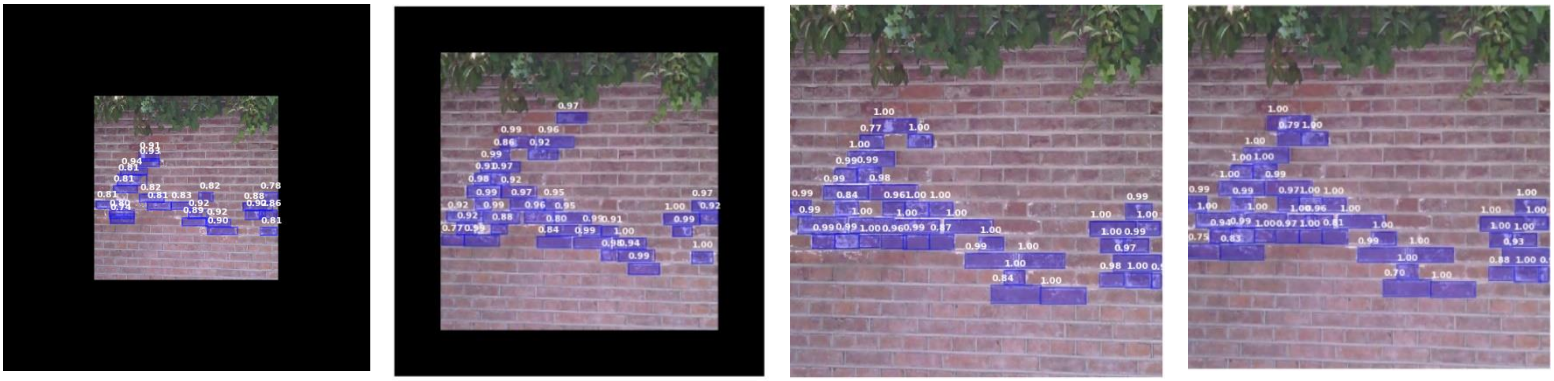


Figure 61 Predicted masks under input scaling (0.5×, 0.75×, 1.25×, 1.5×) compared to the baseline.

RESOLUTION

Reducing image resolution to 320 and 160 pixels substantially degrades visual performance. At 160px, predictions are patchy, misaligned, or completely missing, with efflorescence often undetected. The 480px variant performs comparably to the original high-resolution image, producing reasonably clean contours. This confirms that fine-grained detail is critical for accurate mask generation, especially for subtle or edge-bound efflorescence patterns

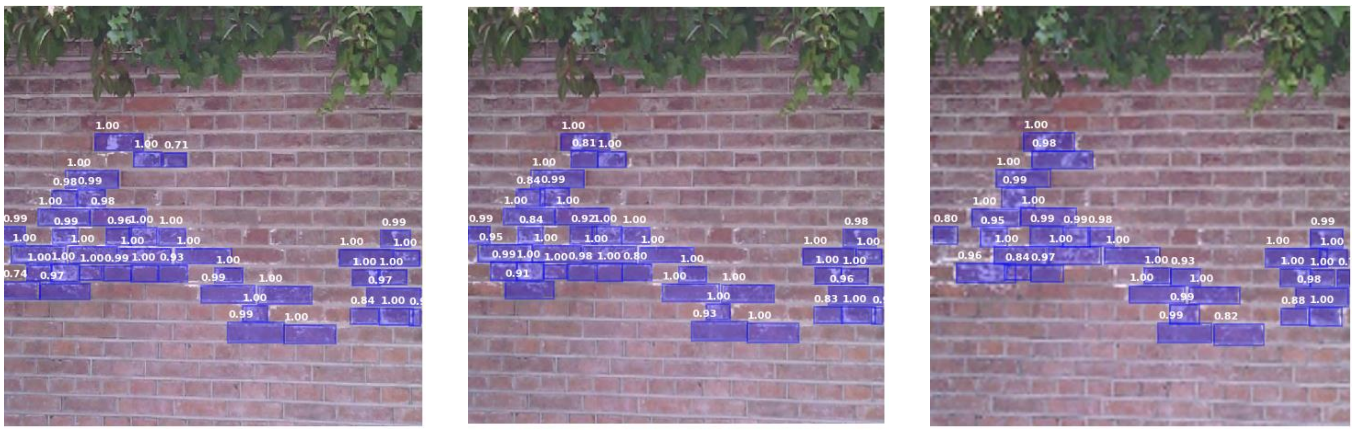


Figure 62 prediction masks at different input resolutions (480px, 320px, 160px) alongside the baseline.

LIGHTING

Predictions on images with moderate brightness increases (+10%) show noticeably improved mask confidence and completeness, with efflorescence outlines clearly defined. In contrast, darker variants (-10%, -20%) suppress mask visibility and produce fragmented or overly conservative detections. The

+20% variant exaggerates boundaries slightly but retains coherent coverage, suggesting the model is more robust to overexposure than underexposure.

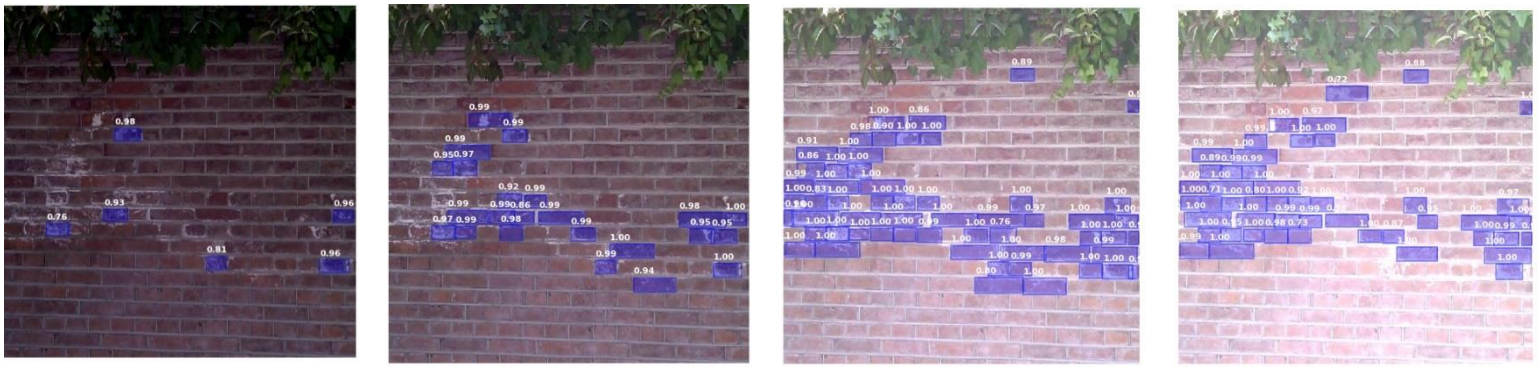


Figure 63Prediction comparisons for brightness/contrast augmentation (±10%, ±20%) versus the baseline

DISCUSSION

By visualizing predicted instance masks across various augmentation scenarios, it becomes clear how specific input affect the model's ability to detect and localize efflorescence on masonry surfaces.

Angular deviations, particularly beyond ±10°, visibly disrupt mask alignment and reduce detection certainty. While the model remains functional under mild rotations, steeper tilts introduce perspective distortion and partial occlusions that impair learning. This suggests that the model lacks sufficient rotational invariance, and that further rotation-aware augmentation strategies or rotation-equivariant architectures could improve robustness under tilted viewpoints.

In terms of scale variation, upscaling (1.25× and 1.5×) leads to denser and more confident predictions, whereas downscaling (especially to 0.5×) degrades both confidence and segmentation accuracy. This is consistent with the understanding that high-resolution inputs retain more texture and edge information, which is essential for precise boundary segmentation. It also highlights a potential challenge for deployment on low-resource devices, where resizing for performance may compromise model accuracy.

Resolution degradation exhibits similar effects: the 160×160 images yield sparse, fragmented predictions with clear losses in detection reliability. Even though training with lower-resolution images is computationally attractive, these findings underscore that efflorescence detection—often characterized by subtle and diffuse patterns—relies heavily on fine visual detail. Therefore, image fidelity remains a critical factor in ensuring robust detection outcomes.

Perhaps the most striking observations arise in the lighting augmentation set. While darker inputs (-20%) lead to visible performance collapse, modest brightness increases (+10%) actually enhance detection. This suggests that the model is sensitive to illumination but has not developed true lighting invariance. Instead, its performance appears tightly coupled to the average luminance and contrast distributions of the training set. These findings imply that future work should prioritize exposure-normalization techniques or training on broader lighting conditions to build a more generalizable model.

Overall, the qualitative evaluation confirms that the model is most vulnerable to extreme cases of downscaling, resolution loss, and lighting degradation. In contrast, it demonstrates moderate tolerance to angular shifts and even benefits from slight exposure enhancements. Together, these results highlight the importance of carefully curating training data to include sufficient variability in scale, lighting, and viewpoint to ensure stable performance in real-world applications.

4.3. Case Study: Real-World Validation

To complement the controlled evaluation of the trained models, a real-world case study was conducted to assess the practical applicability and robustness of the enhanced detection framework. Although the dataset used throughout this research consists entirely of real masonry surfaces photographed under authentic environmental conditions, its images were still limited to predefined experimental setups and annotated samples. The case study therefore extends the validation process by testing the model on an unseen physical site, introducing uncontrolled variables such as lighting, surface irregularities, material heterogeneity, and environmental exposure.

4.3.1. Case Study Context

The selected case study is located at Hodshon-Dedelhof, situated along the Eerste Weteringdwarsstraat near the Vijzelgracht in Amsterdam. The Hodshon-Dedelhof was established in 1842 by Isaac Hodshon (1772–1855) and Isabella Dedel (1778–1865) as a *hofje* (almshouse) intended to provide accommodation for women, particularly those who had served in domestic employment. The courtyard complex remains inhabited exclusively by women and retains much of its original architectural character despite several renovation phases.

Figure 64 Location of the Hodshon-Dedelhof in Amsterdam, The Netherlands

The masonry section selected for this study is a penant (structural pier) forming part of the façade along the Weteringdwarsstraat. The wall measures approximately 28.15 m in length and 2.60 m in height, serving as a characteristic vertical element that separates façade openings and supports the overall façade composition. The wall was renovated approximately five years ago, during which the masonry joints were refilled using a convex pointing profile (bolle voeg) to improve moisture resistance and visual uniformity.

Figure 65 Aerial view of the Hodshon-Dedelhof courtyard complex.

The façade is constructed from red hand-formed clay bricks typical of mid-19th-century Amsterdam architecture, generally measuring around 210 × 100 × 70 mm, laid in a stretcher bond with lime-cement mortar. The renewed joint finish contrasts slightly with the original lime pointing but was applied to maintain consistency across the restored elevation.

A distinctive feature of the penant is the visible salt-related surface deterioration. The wall exhibits efflorescence and sub-florescence patterns distributed in undulating vertical bands, particularly concentrated near the lower sections. These patterns suggest rising damp and leakage phenomena, likely caused by capillary water transport and insufficient drainage at the wall base. The combination of moisture ingress, trapped salts, and a dense repointing layer provides an ideal real-world test case for evaluating the model's ability to detect efflorescence under complex environmental and material conditions.

Figure 66 Orthographic elevation of the selected façade section (penant) along the Eerste Weteringdwarsstraat.

4.3.2. Methodology and Model Setup

The case study was designed to validate the enhanced deep learning models on a real-world masonry surface by integrating 3D reconstruction, localized image extraction, and segmentation-based detection. The workflow mirrors the experimental setup described in Chapter 3 but applies it to an in-situ wall of the Hodshon-Dedelhof.

The wall was recorded using an iPhone equipped with the Polycam LiDAR application, which enabled the generation of a dense 3D mesh and corresponding point cloud with photogrammetric color information. The complete façade section was captured in two separate scans, each consisting of approximately 230 frames and 10 million points. The scans were aligned using Iterative Closest Point (ICP) registration, achieving a final alignment accuracy of $RMSE \approx 0.035\ m$ with a total model extent of $\approx 30.24\ m$.

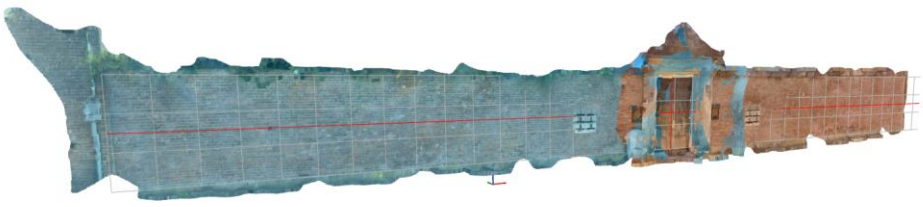


Figure 67 Visualization shows the two scans blue (left) and orange (right) Registered point cloud after ICP with uniform grid acros wall surface (0.75 x 0.75m)

Following alignment, the combined wall contained approximately 21.6 million points. A uniform grid was projected across the registered point cloud to facilitate localized analysis. Based on the findings

of Hypothesis 4, the optimal spatial resolution was set to 640 × 640 pixels, corresponding to an on-site coverage of approximately 0.75 × 0.75 m per tile. This subdivision resulted in 36 images per section (108 in total), from which non-masonry areas such as doors, vegetation, or reflections were excluded.

For model inference, two previously trained detection models were applied:

- the Efflorescence + Damage model, optimized to differentiate between salt crystallization and individual damaged masonry bricks
- the Graffiti + Efflorescence model, used to evaluate robustness against misclassification caused by visually similar surface features.

Each selected tile was processed individually, and the resulting COCO-formatted prediction masks were linked back to their corresponding spatial coordinates through pixel-based mapping. The inference results were then re-projected and overlaid onto the original 3D PLY wall model, preserving their spatial accuracy. In the visualization, blue denotes efflorescence detection, whereas red indicates detected surface damage.

4.3.3. Results and Observations

The comparative visualizations across the three grid configurations illustrate the sensitivity of detection performance to spatial resolution and image scale.

TRIPLE GRID

In the triple-grid configuration (V1), the segmentation produces detailed coverage but with generally low confidence levels for efflorescence detection in both models. While efflorescence regions are still recognized, the predictions appear fragmented and uncertain. Graffiti, on the other hand, is largely undetected, likely due to the smaller image scale per grid cell and the reduced image quality. The model's ability to capture large visual patterns, such as letters or continuous surface discolorations, appears constrained at this resolution.

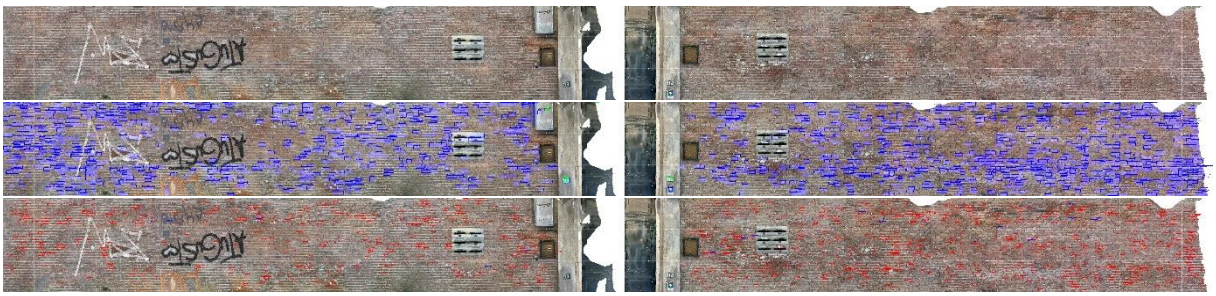


Figure 68 Wall segmentation and detection results Normal (top), Efflorescence and Graffiti (middle), Damage and Efflorescence (bottom) using the base grid (0.75 m × 0.75 m per cell) total of 111 images.

DOUBLE GRID

The double-grid configuration (V2) shows a modest improvement in graffiti recognition, with several graffiti patches successfully detected. However, this setup also introduces false positives, particularly on non-masonry elements such as utility boxes and house numbers, where textural contrasts resemble graffiti. Efflorescence detection confidence slightly decreases at this scale, while damage detection improves significantly, capturing a broader set of local surface irregularities that were previously overlooked in the finer grid.

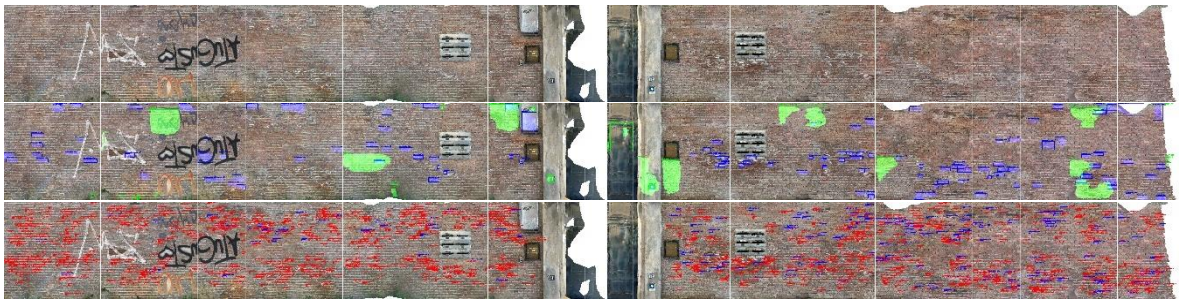


Figure 69 Wall segmentation and detection results Normal (top), Efflorescence and Graffiti (middle), Damage and Efflorescence (bottom) using the base grid (1.125 m × 1.125 m per cell) total of 50 images.

SINGLE GRID

The single-grid configuration delivers the highest overall confidence for efflorescence detection, with large contiguous areas of salt deposition correctly segmented. Graffiti is more consistently detected with higher confidence scores, though false positives remain in regions of poor image quality or uneven lighting. Notably, some bright graffiti marks are misclassified as efflorescence, suggesting that the model remains sensitive to high reflectance or white pigment tones. This scale provides the most stable inference behavior across models but still highlights the limitations introduced by varying surface conditions and capture quality

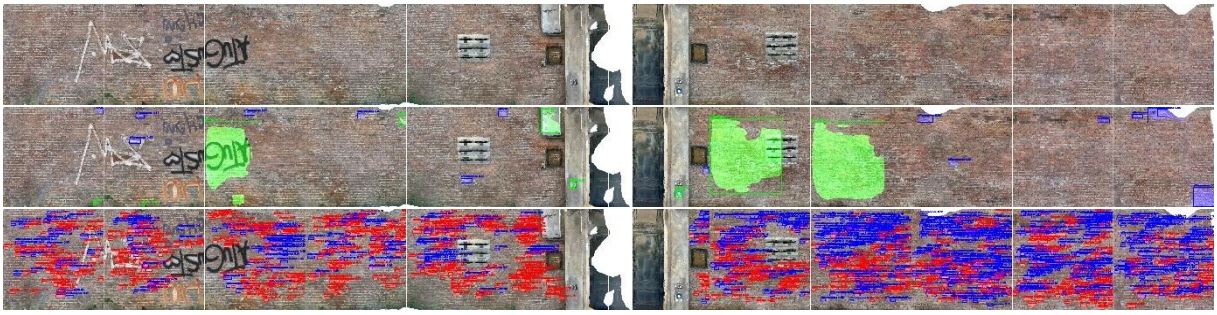


Figure 70 Wall segmentation and detection results Normal (top), Efflorescence and Graffiti (middle), Damage and Efflorescence (bottom) using the base grid (2.5 m × 2.5 m per cell) total of 12 images.

A notable limitation of this case study lies in the relatively low visual quality of the generated image tiles. The reconstruction process constrains mesh resolution, file size, and point density, which in turn reduces the sharpness and fidelity of the extracted textures. Consequently, fine-grained surface features such as micro-cracking or subtle efflorescence boundaries may not be accurately detected by the models, potentially limiting segmentation precision and classification reliability.

5. Discussion

This research has demonstrated the potential of deep learning-based instance segmentation methods, particularly Mask R-CNN and YOLOv8, for detecting masonry damage and efflorescence in heritage and modern buildings. While the results are promising, several technical, methodological, and practical challenges were encountered that impact both model performance and the scalability of this approach.

5.1. Technical Limitations and Model Performance

All models were trained using a ResNet-50 backbone due to frequent out-of-memory (OOM) issues. More powerful hardware could enable the use of deeper backbones (e.g., ResNet-101) or larger batch sizes, potentially improving both convergence and accuracy. Furthermore, training was limited in epochs due to computational constraints, which may have affected final performance.

Transfer learning proved to be a useful strategy. For example, the thermal (RGB-T) model benefited from pre-training on an RGB-only benchmark model for efflorescence detection. This slight performance increase indicates that future work should explore more extensive transfer learning setups, especially when dealing with limited thermal datasets.

The comparison between Mask R-CNN and YOLOv8 highlights the trade-off between accuracy and inference speed. While YOLOv8 is faster, Mask R-CNN offers finer segmentation capabilities. Future research could extend this comparison to newer YOLO models or hybrid architectures.

5.2. Dataset Challenges and Annotation

A major limitation in this study was the diversity and imbalance of the dataset. Unlike prior research that used data from sites with uniform brick types and consistent construction materials, the dataset in this study contains a wide variety of brick textures, colors, and joint styles, especially reflective of modern Dutch buildings. This variability makes detection more difficult. A potential solution is to first classify the type of brick or construction material, allowing for model fine-tuning per subgroup. For instance, the development of efflorescence can vary across materials and textures, as their porosity and composition affect how salts migrate and crystallize..

High-quality annotation was one of the most time-consuming aspects of the research. Brick-by-brick mask annotations are particularly labor-intensive, but essential for accurate detection. While necessary for this study, such annotation processes are difficult to scale. Future work may explore outsourcing or semi-automated annotation pipelines, particularly if larger datasets are to be developed. It should also be noted that mask-based labeling requires significantly more time than bounding box annotation.

Additionally, the model could be extended with a separate class for efflorescence in joints, as this was commonly observed but not explicitly labeled in the current dataset.

5.3. Enhancing Spatial and Contextual Understanding

Spatial context plays a important role in damage detection. As shown in the co-occurrence analysis, damage and efflorescence are more likely to occur in spatial proximity. Including spatial features, such as building location (e.g., coastal proximity, groundwater levels, or orientation) may provide valuable context for improving model performance. Although salts can originate from various sources, integrating environmental data such as mapped groundwater levels could enhance the interpretation of efflorescence occurrence.

Surveyors and practitioners could also contribute to better dataset consistency by following fixed imaging protocols, such as capturing images from standardized distances and angles. This would enhance data usability and reduce noise caused by inconsistent capture practices.

Moreover, incorporating a fifth dimension, such as depth from point clouds or dense imaging layers, could improve the detection of material damage. Damage like disintegration, detachment, or layering is often accompanied by surface depth variation, which RGB imagery alone may fail to capture.

5.4. Future Directions and Recommendations

- Differentiate between damage classes: Current models treat all forms of damage (e.g., scaling, disintegration, blistering) as a single class. Future models should be trained with subclass labels aligned with the MDCS to better capture the nuance of different decay patterns.
- Handle visually similar damage types: Efflorescence often visually resembles lichens, salt crusts, or surface disintegration. Enlarging the dataset and including such confounding examples can improve model robustness and reduce misclassifications.
- Extend to point cloud processing: The ability to process and analyze 3D data (e.g., LiDAR or photogrammetric point clouds) would significantly improve the applicability of this research to heritage conservation and BIM-integrated workflows.
- Expand model selection: Time constraints limited the current model choices to those supported by available research and prior experience. Future work should experiment with transformer-based models, diffusion models, or graph neural networks designed for spatial structures.
- Thermal image enhancement: Additional research is encouraged into thermal differentiation between rising damp and leakage, which may exhibit distinct thermal patterns. Multi-modal fusion techniques could further improve detection accuracy.
- Promote data sharing: A key limitation in the field remains the lack of open annotated datasets. This research addresses that by sharing all data and code openly, encouraging reproducibility and further collaboration.
- While the chi-square analysis of spatial co-occurrence between efflorescence and damage is included in this research, confidence calibration of the model outputs was considered but falls outside the present scope. Future work could explore probability calibration methods, such as logistic regression or post-processing adjustments, to integrate contextual information (e.g., nearby damage) into efflorescence confidence scores.

5.5. Broader Impact

In addition to the scientific contributions, this research responds to real-world challenges in construction monitoring and heritage conservation. The lack of skilled workers, combined with growing maintenance needs, emphasizes the urgency for automated inspection tools. By identifying both visible surface degradation and underlying efflorescence early, such systems can support preventive maintenance strategies, reducing long-term restoration costs.

6. Conclusion

This research explored how deep learning models can be applied to improve the detection of efflorescence in masonry buildings in the Netherlands, with a focus on real-world variability in materials, environmental exposure, and damage appearance. This thesis explored the central research question:

"How can deep learning models be applied to improve the detection of efflorescence in masonry buildings in the Netherlands?"

The study integrated dataset development, model training, hypothesis testing, and spatial analysis to assess the potential and challenges of this approach.

This study demonstrates that deep learning models, particularly instance segmentation frameworks like Mask R-CNN, can be effectively trained to detect efflorescence on masonry surfaces, provided that sufficient data diversity, annotation quality, and spatial understanding are present. Efflorescence is a complex phenomenon, affected by material properties, environmental exposure, and moisture dynamics. Its appearance varies in texture, location (brick face vs. mortar joint), and intensity, which challenges traditional computer vision methods. Deep learning, especially models capable of pixel-wise segmentation, offers the ability to learn subtle visual cues and adapt to varying contexts.

In practice, the application of these models requires several key components. First, the dataset must include a broad range of annotations, surfaces, and damage types to account for the material diversity in Dutch masonry, particularly in urban and historic settings. This research showed that pretraining on generic datasets and transfer learning can be leveraged to overcome initial data scarcity, but model performance improves significantly with tailored, high-quality annotations, especially when bricks are labeled individually with clear damage type definitions and limited diversity in the dataset.

Despite these challenges, the study concludes that deep learning models are a valuable tool for automating efflorescence detection. When combined with carefully designed experiments, consistent annotation standards, and statistical validation, they can support more reliable and scalable masonry inspection workflows. The insights gained here could contribute to preventive maintenance, heritage conservation, and broader efforts to digitize building diagnostics in the Netherlands and beyond.

6.1. Summary of Findings Per Sub-Question

SQ1: What are the visual characteristics of efflorescence on masonry, and how do these factors present challenges for detection?

Efflorescence presents distinct visual characteristics that make it both a valuable indicator of underlying salt-related damage and a difficult target for automated detection. It typically appears as a white crystalline deposit on the surface of brick or mortar, caused by the evaporation of salt related moisture. However, its detectability is highly variable. When masonry is wet, efflorescence may dissolve, temporarily disappearing from view. This solubility makes its appearance intermittent and dependent on environmental conditions such as recent rainfall, humidity, and temperature. As a result, any detection system relying on visual inputs, especially those trained on still images, must contend with the instability of the target damage.

In the context of Dutch masonry, the diversity in brick types and surface textures further complicates visual assessment. Moreover, rising damp and leakage can introduce salts in different ways, leading to highly localized and non-uniform occurrences. Efflorescence often appears near gutters, at the base

of walls, or in corners where moisture movement is pronounced, making it harder to capture with generic imaging approaches.

Another key challenge stems from the misclassification of visually similar phenomena. Surface features such as calcium-based encrustations, biological growth (e.g., lichens, molds), and even graffiti can share textural and tonal similarities with efflorescence, particularly in low-light or overexposed images. These overlapping visual features contribute to false positives in image-based models. Additionally, efflorescence may coexist with other forms of salt-related damage, including alveolization, erosion, delamination, exfoliation, powdering, spalling, crumbling, and blistering, each varying in severity across individual bricks or wall sections. In many heritage buildings, especially those lacking horizontal moisture barriers, salts migrate over time and precipitate at different heights depending on their solubility, complicating consistent annotation and predictions. Such variability reduces annotation consistency and introduces additional noise in the training data, both of which lower detection accuracy.

From a diagnostics perspective, the MDCS framework and earlier foundational work by van Hees and van Balen emphasize that efflorescence is both a symptom of salt and moisture transport processes and an indicator of underlying deterioration mechanisms. While it does not always indicate material damage, it does signal active salt transport within the wall, which can eventually lead to deeper deterioration. Efflorescence can occur either on the surface of the masonry, where it is visible and often considered cosmetic, or within the pores of the brick. Internal crystallization, also known as crypto-efflorescence, exerts pressure on the material from within, potentially leading to mechanical damage such as scaling, spalling, or disintegration. This dual manifestation complicates visual interpretation, as surface efflorescence may appear mild while masking more severe internal deterioration. Consequently, it becomes difficult to determine the exact location and severity of salt-related damage based on surface appearance alone. Heritage experts often rely on contextual information, such as moisture sources, salt types, and environmental data, combined with visual inspection to make informed diagnoses. These nuanced assessments present a challenge for image-based machine learning models, which lack access to subsurface indicators and can only infer damage severity from surface-level features.

These findings underscore that detecting efflorescence is not merely a matter of classification accuracy. It involves coping with material heterogeneity, variable visibility, and context-dependent interpretation. Therefore, machine learning models developed in this study needed to be carefully trained with annotated datasets that account for these complexities—highlighting the importance of domain-specific knowledge in designing both the annotation schema and the detection approach.

SQ2: Which deep learning models are most suitable for detecting and classifying efflorescence on masonry, based on performance criteria?

This study benchmarked two widely adopted deep learning models — Mask R-CNN and YOLOv8 — to evaluate their suitability for detecting efflorescence in heritage masonry. These models represent two fundamentally different object detection paradigms: a two-stage, segmentation-oriented framework (Mask R-CNN) and a one-stage, real-time detection model (YOLOv8). The evaluation was grounded in both theoretical architecture (as detailed in Chapter 2.2) and experimental results across multiple training setups.

Mask R-CNN builds on the Faster R-CNN pipeline, adding a segmentation branch and using a Feature Pyramid Network (FPN) to effectively detect objects across multiple scales. The inclusion of RoI Align preserves pixel-level precision, a critical factor for identifying irregular, diffuse damage patterns such

as efflorescence. Despite higher computational requirements, this architecture supports detailed instance segmentation, making it well-suited for fine-grained masonry analysis.

In contrast, YOLOv8 is designed for speed. As a one-stage, anchor-free detector, it performs classification and localization in a single forward pass. YOLOv8 uses C2f blocks and a multi-scale feature fusion "neck" (similar to FPN) for efficient feature propagation, and outputs predictions directly from its detection head. While this makes YOLOv8 fast and lightweight, it does so at the cost of segmentation quality and spatial granularity.

The models were evaluated using common performance metrics:

- Mean Average Precision (mAP@0.5): Both models reached a modest score of ~0.35, which suggests dataset limitations (small size, inconsistent annotations, varying lighting) rather than model underperformance. However, Mask R-CNN consistently outperformed YOLOv8 in mAP during early and peak epochs.
- Precision: YOLOv8 demonstrated slightly higher precision (0.60) across epochs, indicating a lower false positive rate. This suggests that YOLOv8 is more selective in assigning detections, which aligns with its bounding-box-first design.
- Recall: YOLOv8 also achieved higher recall, meaning it was better at detecting all instances of efflorescence, albeit sometimes at the cost of overgeneralization or false positives.
- Loss Trends: Mask R-CNN exhibited lower and more stable classification and bounding box regression loss, indicating stronger learning stability and more precise region proposals.
- Inference Speed: YOLOv8's anchor-free design and single-pass architecture resulted in higher theoretical throughput, though in this setup.

Qualitative visualizations revealed clear behavioral differences. Mask R-CNN produced detailed masks with tight boundary alignment, making it ideal for pixel-level mapping. However, it sometimes over-segmented or misclassified faint discolorations. YOLOv8, while unable to provide masks, offered broader coverage with dense bounding boxes, which occasionally resulted in less spatial precision.

Misclassification between efflorescence and visually similar damages (e.g., encrustation, lichens, graffiti) occurred in both models, though Mask R-CNN appeared more prone to false positives, while YOLOv8 missed finer detections. These results underscore the challenge of visually ambiguous features and the importance of dataset refinement.

Despite the advantages of YOLOv8 in speed and recall, the spatial precision, segmentation quality, and stability of Mask R-CNN make it more suitable for the fine-grained detection tasks required in heritage masonry studies. Especially when understanding the extent, shape, and overlap of damage is important, such as differentiating efflorescence from adjacent surface deterioration, Mask R-CNN offers superior utility.

However, both models are affected by the underlying dataset variability, not just architecture. Low mAP and limited generalization highlight the need for future improvements in data annotation consistency, lighting normalization, and scale standardization, all of which would enhance learning and prediction stability.

SQ3: What is the effect of variables (such as image quality, lighting, and orientation) on the performance of the model?

The third sub-question addresses one of the most critical practical challenges in deploying deep learning for efflorescence detection: how environmental and visual acquisition factors affect model accuracy. Field testing and dataset inspection revealed that real-world variation in image quality, lighting, scale, and angle introduces significant performance degradation, particularly in cases where efflorescence appears faint, fragmented, or unevenly illuminated. These inconsistencies often result in segmentation artifacts or missed detections, especially on reflective, dark, or shaded surfaces.

To address these challenges, this study explored both annotation strategy and augmentation design. A pivotal decision was the transition from class-agnostic efflorescence masks to brick-level annotations. This revised strategy helped the model contextualize damage in relation to the masonry unit, thereby improving spatial relevance and interpretability, especially when assessing location-based causes like rising damp or material degradation. Previous attempts at multi-class annotations (e.g., distinguishing between efflorescence on mortar joints versus bricks) were abandoned due to class imbalance and limited training benefit.

In parallel, augmentation pipelines were developed using OpenCV and Albumentations to simulate real-world variability. Transformations included controlled adjustments in brightness, contrast, resolution, angle, and scale. These augmentations were essential for testing Hypothesis 4, which posits that increasing dataset diversity improves model robustness under non-ideal conditions.

Literature highlights the importance of such augmentation: studies that achieve high precision often use controlled orthogonal datasets, whereas field-acquired datasets like in this study, with varying exposure, framing, and resolution, present greater learning complexity. This variation was evident in image resolution (from 232×300 to 5858×3911 px) and acquisition devices (smartphones vs. DSLR), which complicates feature consistency and affects generalization.

The results of the augmentation experiments confirm that model performance is indeed sensitive to changes in image quality, viewpoint, and lighting. Quantitative analyses showed that mild perturbations, such as ±5° rotations or modest scale increase, had limited effect on performance. However, more extreme transformations, particularly image downscaling (0.5×), severe brightness reductions (-20%), and low resolution inputs (160×160), resulted in substantial drops in detection accuracy and unstable training behavior.

Qualitative results reinforced these observations: under low-light conditions or reduced image resolution, the model frequently failed to detect efflorescence or produced fragmented masks with low confidence. In contrast, certain augmentations, like slight overexposure (+10% brightness), even improved detection performance, suggesting the model benefits from well-lit, high-contrast inputs. Angular deviations beyond ±10° and strong resolution loss were particularly detrimental, leading to distorted perspective and blurred texture detail that interfered with spatial recognition.

These findings support the hypothesis that strategic augmentation can help mitigate sensitivity to acquisition variability. However, they also underscore the limitations of current architectures when faced with domain shift. The study shows that while augmentations improve general robustness, they do not fully compensate for the underlying dependence on visual clarity and consistent framing, especially for detecting subtle, diffuse damage patterns like efflorescence.

SQ4: How can the model performance be improved by addressing misclassification of similar damage types and co-occurrence with efflorescence?

This sub-question aimed to tackle two persistent challenges in efflorescence detection: (1) misclassification with visually similar surface features such as encrustation, lichens, and graffiti, and (2) the difficulty in detecting co-occurring damage types within the same or neighbouring masonry units.

Initial evaluations of Mask R-CNN and YOLOv8 models revealed a clear issue: misclassification of efflorescence, especially when confused with similar-looking damage types like encrustation, graffiti, and biological growths. These errors arose from overlapping visual cues, such as white surface patches, irregular textures, or environmental staining, which challenged not only the model but also human observers.

To mitigate this, the models were retrained using a multi-class strategy, assigning separate categories to efflorescence, encrustation, lichens, and graffiti. The following patterns emerged:

- Graffiti reached the highest mAP (0.6) and precision near 1.0, attributed to its strong color contrast with masonry. As seen in the confusion matrix, it had 67 true positives, only 1 false positive, and 2 false negatives, reflecting stable detection with minimal confusion.
- Lichens emerged as the most robust new class, showing perfect classification by epoch 60: 29 true positives, no false positives or false negatives. This was supported by mAP curves and high precision/recall, indicating the model learned to distinguish lichens based on their unique texture and color.
- Encrustation, however, exhibited higher segmentation uncertainty: while true positives were comparable (36), it also produced 5 false positives and 4 false negatives, indicating that the model often over-segmented or falsely triggered on rough surfaces. mAP fluctuated between 0.35 and 0.5 across epochs, and its recall and precision lagged behind other classes.
- Efflorescence, while achieving consistent mask shapes and relatively high recall (0.75–0.8 in some epochs), showed the most instability in precision and mAP. For example, in one test case, it achieved 33 true positives, but still had 1–2 false positives depending on the pairwise test. This instability is further confirmed by the fluctuating mAP and recall curves over time.

These results, illustrated clearly in your evaluation plots and confusion matrices, underscore the importance of visual distinctiveness and annotation clarity. Where graffiti and lichens are distinct in texture or hue, efflorescence's diffuse appearance and overlap with damage artifacts like encrustation lead to confusion. Nonetheless, the multi-class setup did help compartmentalize class-specific features, improving detection consistency compared to the original single-class model.

The final improvement was evaluated through a co-occurrence hypothesis, assessing whether damage frequently occurs near efflorescence. A Chi-square test based on centroid analysis of ground truth annotations confirmed a statistically significant correlation ($p < 0.00001$) between efflorescence and adjacent damage, especially within 1–2× brick-width range. However, the model itself often failed to detect both classes on the same brick, motivating the use of annotations over predictions for co-occurrence testing. This highlights an architectural limitation in the model's ability to recognize overlapping or co-existing damage features.

To explore whether damage frequently occurs near efflorescence, a Mask R-CNN model was trained to detect both classes on annotated masonry images. The model was trained for 60 epochs using a dataset annotated for both efflorescence and physical damage (e.g., disintegration, loss of cohesion, layering). Training progress was monitored via total loss, classification loss, bounding box loss, and mask loss. As seen in the logs, total loss dropped sharply in the early epochs and plateaued around

1.0 after epoch 30, indicating basic convergence. The bounding box loss converged rapidly (<0.1), suggesting reasonable spatial localization. However, classification loss remained relatively high (0.5), indicating difficulty in distinguishing between damage and efflorescence, especially when both appear in the same region. Mask loss stabilized around 0.15–0.2, reflecting moderate segmentation performance but leaving room for improvement.

Evaluation results highlighted further limitations. Class-specific recall for the combined detection of efflorescence and damage increased slowly, reaching only 0.40 by epoch 60, with significant fluctuations. Precision started high (0.75) but declined steadily, stabilizing around 0.55–0.60, reflecting a growing rate of false positives during training. Meanwhile, mAP@0.5 performance was stronger for damage (0.45) than efflorescence (0.30–0.35), likely due to efflorescence's more diffuse and ambiguous appearance. Notably, both metrics showed instability across epochs, with sharp spikes and drops, which undermines model reliability. These fluctuations are likely due to dataset imbalances, noisy or overlapping annotations, and the challenge of detecting co-located phenomena.

Crucially, qualitative inspection revealed that the model frequently failed to detect both classes within the same brick, even when the annotations clearly indicated overlapping efflorescence and damage. Instead, the model tended to isolate only one class, missing dual-presence cases entirely. This architectural limitation in instance-based segmentation, combined with training instability, led to the decision to base the spatial co-occurrence analysis on ground truth annotations rather than model predictions.

For the statistical evaluation, a Chi-square test of independence was performed on centroid-based ground truth annotations to test the hypothesis:

- $\mathbf{H_0}$ (Null Hypothesis): The presence of damage in a brick is independent of efflorescence in its neighboring bricks.
- $\mathbf{H_1}$ (Alternative Hypothesis): Bricks near efflorescence are more likely to exhibit damage.

Annotations were analyzed using their centroid positions, with proximity zones defined relative to average brick width (e.g., within 1×, 2×, and 5× distances). The results confirmed a significant spatial association between efflorescence and nearby damage ($p < 0.00001$ for 1–2× width zones), strongly supporting H_1. Although the model could not reliably capture dual-label instances, the annotated data clearly showed that these forms of degradation tend to co-occur, especially within close spatial proximity. Incorporating additional data, and running interference by different single class models could increase the detection accuracy by dual-label instances overlapping as a post processing step.

SQ5: How can the integration of thermal (IR) imagery improve the detection accuracy and reliability of efflorescence in masonry?

To evaluate Hypothesis 1, whether thermal imaging improves efflorescence detection in moisture-related masonry, an RGB-Thermal (RGBT) model was developed and compared to a baseline RGB-only setup. By incorporating thermal data as a fourth channel, the model aimed to capture moisture gradients not always visible in RGB images, particularly in cases involving rising damp, leakage, or salt accumulation. The RGBT model was trained on 150 spatially aligned RGB and thermal image pairs, with the goal of increasing detection reliability under ambiguous conditions.

Quantitative results confirmed the benefits of thermal fusion in reducing false positives and boosting model confidence. The RGBT model consistently achieved higher precision (up to 0.94) and lower false positive counts (as low as 3 per evaluation set), reflecting more selective and reliable segmentation. It also maintained a high average confidence (0.96), suggesting that the predictions it

made were typically well-founded. In contrast, the RGB-only model leaned toward high recall (up to 0.96) and mAP (up to 0.84), but suffered from a greater number of false positives (up to 18), especially in scenes with texture-heavy or light-stained surfaces. Both models ultimately reached comparable F1-scores (0.76–0.84), though through different strategies: RGB favored exhaustive coverage, while RGBT prioritized high-certainty detections.

However, the RGBT model introduced new challenges during training. Its loss and mAP curves were more erratic, particularly in early epochs, indicating sensitivity to the added modality and a higher risk of underfitting or over-regularization. The RGB model, in contrast, showed smoother convergence and lower total loss across class, mask, and box components. Additionally, the RGBT model tended to miss valid instances, sometimes yielding over 20 false negatives, suggesting it was more conservative when thermal data did not strongly indicate moisture presence. These issues point to the need for longer training cycles and more diverse input to stabilize the benefits of thermal fusion.

Despite these trade-offs, the RGBT model proved more reliable in field-like conditions where false positives are harder to filter manually, such as historic façades or poorly lit interiors. Its ability to avoid spurious detections makes it particularly suited for applications where over-segmentation is costly, including automated damage mapping or restoration prioritization. Overall, the hypothesis is supported: thermal imaging improves detection reliability in moisture-related cases, but also requires careful training and use-case-specific balancing between recall and certainty.

SQ6: How well does the enhanced model perform when evaluated on unseen data and applied to real-world case studies of efflorescence?

When applied to the real-world case study, the enhanced model demonstrated a consistent ability to identify efflorescence across varying grid configurations, though overall performance was influenced by spatial scale, image quality, and surface complexity. The triple-grid configuration (V1) produced detailed but fragmented segmentation, characterized by low confidence scores for efflorescence and limited graffiti recognition. The double-grid configuration (V2) improved graffiti detection and local surface response but introduced false positives on non-masonry features. The single-grid configuration yielded the most stable results, with high efflorescence confidence and improved damage consistency, though bright graffiti regions were occasionally misclassified as efflorescence.

Despite these outcomes, several practical limitations remain when applying the models to real-world environments. The dataset and field-acquired imagery contain substantial variation in resolution, scale, and lighting, leading to inconsistent feature representation between samples. This diversity mirrors real-world heterogeneity but complicates precise segmentation under uncontrolled capture conditions. Differences in camera type, framing distance, and occlusion (e.g., vegetation, surface reflections, or construction elements) further challenge the model's robustness and generalization ability. Moreover, the image reconstruction process used for case study visualization constrains mesh resolution, file size, and point density, reducing texture sharpness and limiting the model's ability to identify fine-grained surface details such as micro-cracks or subtle efflorescence boundaries. These limitations illustrate the broader challenges of scaling deep learning models from controlled training datasets to in-situ heritage environments, where variations in geometry, lighting, and material weathering are unavoidable.

Nevertheless, the model's capacity to reliably detect efflorescence under diverse, unseen conditions demonstrates potential for practical diagnostic use. Future improvements in dataset quality, camera calibration, and high-resolution reconstruction workflows could further enhance segmentation precision and make such models more deployable for large-scale, on-site heritage monitoring.

References

Alexakis, E., Delegou, E., Mavrepis, P., … A. R.-C. S. in, & 2024, undefined. (2024). A novel application of deep learning approach over IRT images for the automated detection of rising damp on historical masonries. *Elsevier*. https://www.sciencedirect.com/science/article/pii/S2214509524000408

Alexakis, E., Delegou, E. T., Mavrepis, P., Rifios, A., Kyriazis, D., & Moropoulou, A. (2024). A novel application of deep learning approach over IRT images for the automated detection of rising damp on historical masonries. *Case Studies in Construction Materials*, *20*, e02889. https://doi.org/10.1016/J.CSCM.2024.E02889

Blauer, C., Kueng, A., Zehnder, K., Böhma, C. B., Künga, A., & Zehnderb, K. (2001). Salt crystal intergrowth in efflorescence on historic buildings. *Researchgate.Net*. https://doi.org/10.2533/chimia.2001.996

Bonduel, M., Klein, R., Vergauwen, M., & Pauwels, P. (2021). *A framework for a linked data-based heritage BIM*. https://lirias.kuleuven.be/3416395&lang=en

Boot, W., Terwel, K., … H. S. the I. of C. E., & 2015, undefined. (2015). Legal matters related to structural damage in the Netherlands. *Icevirtuallibrary.ComW Boot, K Terwel, H StrangProceedings of the Institution of Civil Engineers-Forensic Engineering, 2015•icevirtuallibrary.Com*, *168*(3), 117–126. https://doi.org/10.1680/feng.14.00017

Brick reader. (2024). *Reader brick_060712*.

Charola, A. E., & Bläuer, C. (2015a). Salts in Masonry: An Overview of the Problem. *Restoration of Buildings and Monuments*, *21*(4–6), 119–135. https://doi.org/10.1515/RBM-2015-1005/PDF

Charola, A. E., & Bläuer, C. (2015b). Salts in Masonry: An Overview of the Problem. *Restoration of Buildings and Monuments*, *21*(4–6), 119–135. https://doi.org/10.1515/RBM-2015-1005/PDF

Davies, O., … W. B.-I. J. of, & 2024, undefined. (2024). Challenges to Implementation of Adaptive Reuse of Heritage Buildings. *Journalspub.ComOOA Davies, WG Brisibe, IEE DaviesInternational Journal of Architectural Heritage, 2024•journalspub.Com*. https://doi.org/10.37628/IJAH

Fino, M. De, Galantucci, R., Heritage, F. F.-, & 2023, undefined. (2023). Condition assessment of heritage buildings via photogrammetry: A scoping review from the perspective of decision makers. *Mdpi.ComM De Fino, RA Galantucci, F FatigusoHeritage, 2023•mdpi.Com*. https://doi.org/10.3390/heritage6110367

Girshick, R., Donahue, J., Darrell, T., & Malik, J. (2016). Region-Based Convolutional Networks for Accurate Object Detection and Segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *38*(1), 142–158. https://doi.org/10.1109/TPAMI.2015.2437384

Gonçalves, A., de Brito, J., & Amaro, B. (2015). Systematic Approach to Inspect, Diagnose, and Repair Masonry Walls. *Journal of Performance of Constructed Facilities*, *29*(6). https://doi.org/10.1061/(ASCE)CF.1943-5509.0000650

Goudeau, J. (2015). *Biografie van de baksteen 1850-2000 Zwolle/Amersfoort: WBooks/Rijksdienst voor het Cultureel Erfgoed, 2012 9789040007569 Historisch metselwerk*. https://doi.org/10.7480/knob.114.2015.1

Hall, C., Baird, T., James, M., Tourism, Y. R.-J. of H., & 2016, undefined. (2016). Climate change and cultural heritage: conservation and heritage tourism in the Anthropocene. *Taylor & Francis*, *11*(1), 10–24. https://doi.org/10.1080/1743873X.2015.1082573

Hatir, M. E., Barstuğan, M., & İnce, İ. (2020). Deep learning-based weathering type recognition in historical stone monuments. *Journal of Cultural Heritage*, *45*, 193–203. https://doi.org/10.1016/J.CULHER.2020.04.008

Hatır, M. E., İnce, İ., & Korkanç, M. (2021). Intelligent detection of deterioration in cultural stone heritage. *Journal of Building Engineering*, *44*, 102690. https://doi.org/10.1016/J.JOBE.2021.102690

Hees, R. van. (1995). *The masonry damage diagnostic system*. https://www.academia.edu/88208707/The_masonry_damage_diagnostic_system

Hees, R. van, Naldini, S., & Lubelli, B. (2009). The development of MDDS-COMPASS. Compatibility of plasters with salt loaded substrates. *Construction and Building Materials*, *23*(5), 1719–1730. https://doi.org/10.1016/J.CONBUILDMAT.2008.08.010

IHBC. (2021). *Retrofitting of Traditional Buildings* .

Jung, J. J., & Mazzetto, S. (2024). Integrating Emerging Technologies with Digital Twins for Heritage Building Conservation: An Interdisciplinary Approach with Expert Insights and Bibliometric Analysis. *Heritage 2024, Vol. 7, Pages 6432-6479*, *7*(11), 6432–6479. https://doi.org/10.3390/HERITAGE7110300

Keshmiry, A., Hassani, S., Dackermann, U., & Li, J. (2024). Assessment, repair, and retrofitting of masonry structures: A comprehensive review. *Construction and Building Materials*, *442*, 137380. https://doi.org/10.1016/J.CONBUILDMAT.2024.137380

Korswagen, P. A., Longo, M., Prosperi, A., Rots, J. G., & Terwel, K. C. (2024). Modelling of Damage in Historical Masonry Façades Subjected to a Combination of Ground Settlement and Vibrations. *RILEM Bookseries*, *47*, 904–917. https://doi.org/10.1007/978-3-031-39603-8_73

Koster, H. R. A., & Rouwendal, J. (2017). Historic amenities and housing externalities: evidence from the Netherlands. *Economic Journal*, *127*(605), F396–F420. https://doi.org/10.1111/ecoj.12477

Labadi, S., Giliberto, F., Rosetti, I., … L. S.-… J. of H., & 2021, undefined. (2021). Heritage and the sustainable development goals: Policy guidance for heritage and development actors. *Kar.Kent.Ac.Uk*. https://kar.kent.ac.uk/89231/1/ICOMOS_SDGs_Policy_Guidance_2021.pdf

Lazrak, F., Nijkamp, P., Rietveld, P., & Rouwendal, J. (2014). The market value of cultural heritage in urban areas: An application of spatial hedonic pricing. *Journal of Geographical Systems*, *16*(1), 89–114. https://doi.org/10.1007/S10109-013-0188-1

LeCun, Y., Bottou, L., Bengio, Y., & Haffner, P. (1998). Gradient-based learning applied to document recognition. *Proceedings of the IEEE*, *86*(11), 2278–2323. https://doi.org/10.1109/5.726791

Lin, T.-Y., Dollar, P., Girshick, R., He, K., Hariharan, B., & Belongie, S. (2017). *Feature Pyramid Networks for Object Detection* (pp. 2117–2125).

Liu, L., Ouyang, W., Wang, X., Fieguth, P., Chen, J., Liu, X., & Pietikäinen, M. (2020). Deep Learning for Generic Object Detection: A Survey. *International Journal of Computer Vision*, *128*(2), 261–318. https://doi.org/10.1007/S11263-019-01247-4/FIGURES/21

Lopez-Arce, P., Doehne, E., Greenshields, J., Benavente, D., & Young, D. (2009). Treatment of rising damp and salt decay: the historic masonry buildings of Adelaide, South Australia. *Materials and Structures/Materiaux et Constructions*, *42*(6), 827–848. https://doi.org/10.1617/S11527-008-9427-1

Lourenço, P., Hees, R. van, … F. F.-S. rehabilitation of, & 2014, undefined. (2014). Characterization and damage of brick masonry. *SpringerPB Lourenço, R van Hees, F Fernandes, B LubelliStructural Rehabilitation of Old Buildings, 2014•Springer*, 109–130. https://doi.org/10.1007/978-3-642-39686-1_4

Lubelli, B., Van Hees, R. P. J., & Groot, C. J. W. P. (2004). The role of sea salts in the occurrence of different damage mechanisms and decay patterns on brick masonry. *Construction and Building Materials*, *18*(2), 119–124. https://doi.org/10.1016/j.conbuildmat.2003.08.017

Makoond, N., & Pela, L. (2021). A risk index for the structural diagnosis of masonry heritage (RISDiMaH). *ElsevierN Makoond, L Pela, C MolinsConstruction and Building Materials, 2021•Elsevier*. https://www.sciencedirect.com/science/article/pii/S0950061821001938

Mansuri, L. E., & Patel, D. A. (2022). Artificial intelligence-based automatic visual inspection system for built heritage. *Smart and Sustainable Built Environment*, *11*(3), 622–646. https://doi.org/10.1108/SASBE-09-2020-0139/FULL/XML

Marín-García, D., Bienvenido-Huertas, D., Carretero-Ayuso, M. J., & Torre, S. Della. (2023). Deep learning model for automated detection of efflorescence and its possible treatment in images of brick facades. *Automation in Construction*, *145*, 104658. https://doi.org/10.1016/J.AUTCON.2022.104658

Michalski, S., Cci, J., & Luiz, P. (2016). *The ABC Method: a risk management approach to the preservation of cultural heritage*. https://www.iccrom.org/sites/default/files/2017-12/risk_manual_2016-eng.pdf

M.S. Bakker, E. Homburg, Dick van Lente, & H.W. Lintsen. (1993). *Geschiedenis van de techniek in Nederland. De wording van een moderne samenleving 1800-1890. Deel III*. https://www.dbnl.org/tekst/lint011gesc03_01/lint011gesc03_01_0014.php

Nijland, T., Lubelli, B., & van Hees. (2018). Een plaag van alle tijden: zout: Over oude en toekomstige schade, oud en toekomstig onderzoek. *Research.Tudelft.Nl*. https://research.tudelft.nl/en/publications/een-plaag-van-alle-tijden-zout-over-oude-en-toekomstige-schade-ou

Pan, H., Azimi, M., Gui, G., Yan, F., & Lin, Z. (2018). Vibration-based support vector machine for structural health monitoring. *Lecture Notes in Civil Engineering*, *5*, 167–178. https://doi.org/10.1007/978-3-319-67443-8_14

Pintossi, N., Kaya, D., Wesemael, P. van, International, A. R.-H., & 2023, undefined. (n.d.). Challenges of cultural heritage adaptive reuse: A stakeholders-based comparative study in

three European cities. *Elsevier*. Retrieved January 3, 2025, from https://www.sciencedirect.com/science/article/pii/S019739752300067X

Proietti, N., Calicchia, P., Colao, F., De Simone, S., Tullio, V. Di, Luvidi, L., Prestileo, F., Romani, M., Tatì, A., Rosina, E., & Lubelli, B. (2021). Moisture damage in ancient masonry: a multidisciplinary approach for in situ diagnostics. *Mdpi.Com*. https://doi.org/10.3390/min11040406

Ren, S., He, K., Girshick, R., & Sun, J. (2015). Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks. *Advances in Neural Information Processing Systems*, *28*. https://github.com/

Roy, D., & Kalidindi, S. N. (2017). Critical challenges in management of heritage conservation projects in India. *Journal of Cultural Heritage Management and Sustainable Development*, *7*(3), 290–307. https://doi.org/10.1108/JCHMSD-03-2017-0012/FULL/HTML

SA Smith. (2011). Duties, liabilities, and damages. *HeinOnline*. https://heinonline.org/hol-cgi-bin/get_pdf.cgi?handle=hein.journals/hlr125&section=91

Sabbioni, C., Brimblecombe, P., & A Bonazza. (2007). Mapping climate change and cultural heritage. *Academia.Edu*. https://www.academia.edu/download/31157702/A-1_182.pdf#page=137

Sabbioni, C., Cassar, M., … P. B.-… and M. M., & 2008, undefined. (2008). Vulnerability of cultural heritage to climate change. *Coe.Int*, 44. http://www.coe.int/t/dg4/majorhazards/activites/2009/ravello15-16may09/Ravello_APCAT2008_44_Sabbioni-Jan09_EN.pdf

Sapkota, R., Ahmed, D., & Karkee, M. (2023). Comparing YOLOv8 and Mask RCNN for object segmentation in complex orchard environments. *Artificial Intelligence in Agriculture*, *13*, 84–99. https://doi.org/10.1016/j.aiia.2024.07.001

Sapkota, R., Ahmed, D., & Karkee, M. (2024). Comparing YOLOv8 and Mask R-CNN for instance segmentation in complex orchard environments. *Artificial Intelligence in Agriculture*, *13*, 84–99. https://doi.org/10.1016/J.AIIA.2024.07.001

Sesana, E., Gagnon, A., Bertolin, C., & J Hughes -. (2018). Adapting cultural heritage to climate change risks: perspectives of cultural heritage experts in Europe. *Mdpi.Com*. https://doi.org/10.3390/geosciences8080305

Soleymani, A., & Jahangir, H. (2023). Damage detection and monitoring in heritage masonry structures: Systematic review. *ElsevierA Soleymani, H Jahangir, ML NehdiConstruction and Building Materials, 2023•Elsevier*. https://www.sciencedirect.com/science/article/pii/S0950061823021189?casa_token=TNtbJEmqEn4AAAAA:iSBimM3eLedMGmrKla1stosYv8_HE5r8JgVVf6RYQobFye3LMndSin-89cXLk251GM8Isc4S

Steiger, M., Charola, A. E. 4, & Sterflinger, K. (2011). *Weathering and Deterioration*. https://doi.org/10.1007/978-3-642-14475-2_4

Strlič, M., Thickett, D., Taylor, J., Conservation, M. C.-S. in, & 2013, undefined. (2013). Damage functions in heritage science. *Taylor & Francis*, *58*(2), 80–87. https://doi.org/10.1179/2047058412Y.0000000073

UNESCO. (2024). *World Heritage Centre - Guidance and Toolkit for Impact Assessments in a World Heritage Context*. https://whc.unesco.org/en/guidance-toolkit-impact-assessments/

Van Balen, K. (1998). *Scientific tools for assessment of the degradation of historic brick masonry*.

van Hees, R. P. J., & Naldini, S. (2020). MDCS - A system for damage identification and monitoring. *Preventive Conservation - From Climate and Damage Monitoring to a Systemic and Integrated Approach - Proceedings of the International WTA - PRECOM3OS Symposium*, 113–118. https://doi.org/10.1201/9781003004042-17

Vandemeulebroucke, I., Kotova, L., Caluwaerts, S., & Van Den Bossche, N. (2023). Degradation of brick masonry walls in Europe and the Mediterranean: Advantages of a response-based analysis to study climate change. *Building and Environment*, *230*, 109963. https://doi.org/10.1016/J.BUILDENV.2022.109963

Wang, N., Zhao, Q., Li, S., … X. Z.-C. C., & 2018, undefined. (2018). Damage classification for masonry historic structures using convolutional neural networks based on still images. *Wiley Online LibraryN Wang, Q Zhao, S Li, X Zhao, P ZhaoComputer-Aided Civil and Infrastructure Engineering, 2018•Wiley Online Library*, *33*(12), 1073–1089. https://doi.org/10.1111/mice.12411

Wang, N., Zhao, X., Zhao, P., Zhang, Y., Zou, Z., & Ou, J. (2019). Automatic damage detection of historic masonry buildings based on mobile deep learning. *Automation in Construction*, *103*, 53–66. https://doi.org/10.1016/J.AUTCON.2019.03.003

World Heritage Centre - World Heritage and Sustainable Development. (n.d.). Retrieved July 16, 2025, from https://whc.unesco.org/en/sustainabledevelopment/

Yu, Y., Abu Raed, A., Peng, Y., Pottgiesser, U., Verbree, E., van Oosterom, P., Khaimah, A., & Khaimah, R. A. (n.d.). How digital technologies have been applied for architectural heritage risk management: a systemic literature review from 2014 to 2024. *Nature.Com*. https://doi.org/10.1038/s40494-025-01558-5

Nypan, T. (2006). Cultural heritage monuments and historic buildings as value generators in a post-industrial economy. With emphasis on exploring the role of the sector as economic driver. Rev.

EC Energy (2022), available at: https://energy.ec.europa.eu/topics/energy-efficiency/energy-efficient-buildings/energy-performance-buildings-directive_en

Sabbioni, C., Cassar, M., Brimblecombe, P., & Lefevre, R. A. (2008). Vulnerability of cultural heritage to climate change. European and Mediterranean Major Hazards Agreement (EUR-OPA), 1-24.

Cultural Heritage Agency of the Netherlands. (n.d.). Masonry. Retrieved January 4, 2025, from https://kennis.cultureelerfgoed.nl/index.php/Baksteenmetselwerk#:~:text=Wind%20en%20water%20leiden%20op,waardoor%20er%20scheuren%20kunnen%20ontstaan.

Snepvangers, K. (2005). Wat, waarom, zinvol meten van bouwfysische elementen. In D. Van Gemert, R. Van Hees, & H. Schellen (Eds.), Monitoring en Diagnose (pp. 34-49). Delft, Nederland: WTA Vlaanderen https://www.wta

international.org/fileadmin/user_upload/Nederland-Vlaanderen/syllabi/oude-syllabi/Monitoring_en_diagnose.pdf

Rijksdienst voor de Monumentenzorg (RDMZ). (2005). Restauratie en beheer (3rd ed.). Utrecht, Nederland: Hoonte Bosch & Keuning

Vanhellemont, Y. (2008). *Vocht door hygroscopische zouten: Wat doe je eraan? Toepassing op (deels) ingegraven metselwerk.* WTA Nederland - Vlaanderen. ISBN 978-90-79216-02-

Jocher, G., Chaurasia, A., & Qiu, J. (2023). *Ultralytics YOLOv8 (Version 8.0.0)* [Computer software]. Ultralytics. https://github.com/ultralytics/ultralytics

Dupont, M. (2024, May 1). *YOLOv8 vs Mask R-CNN: In-depth analysis and comparison*. Retrieved from https://www.labelvisor.com/yolov8-vs-mask-r-cnn-in-depth-analysis-and-comparison/

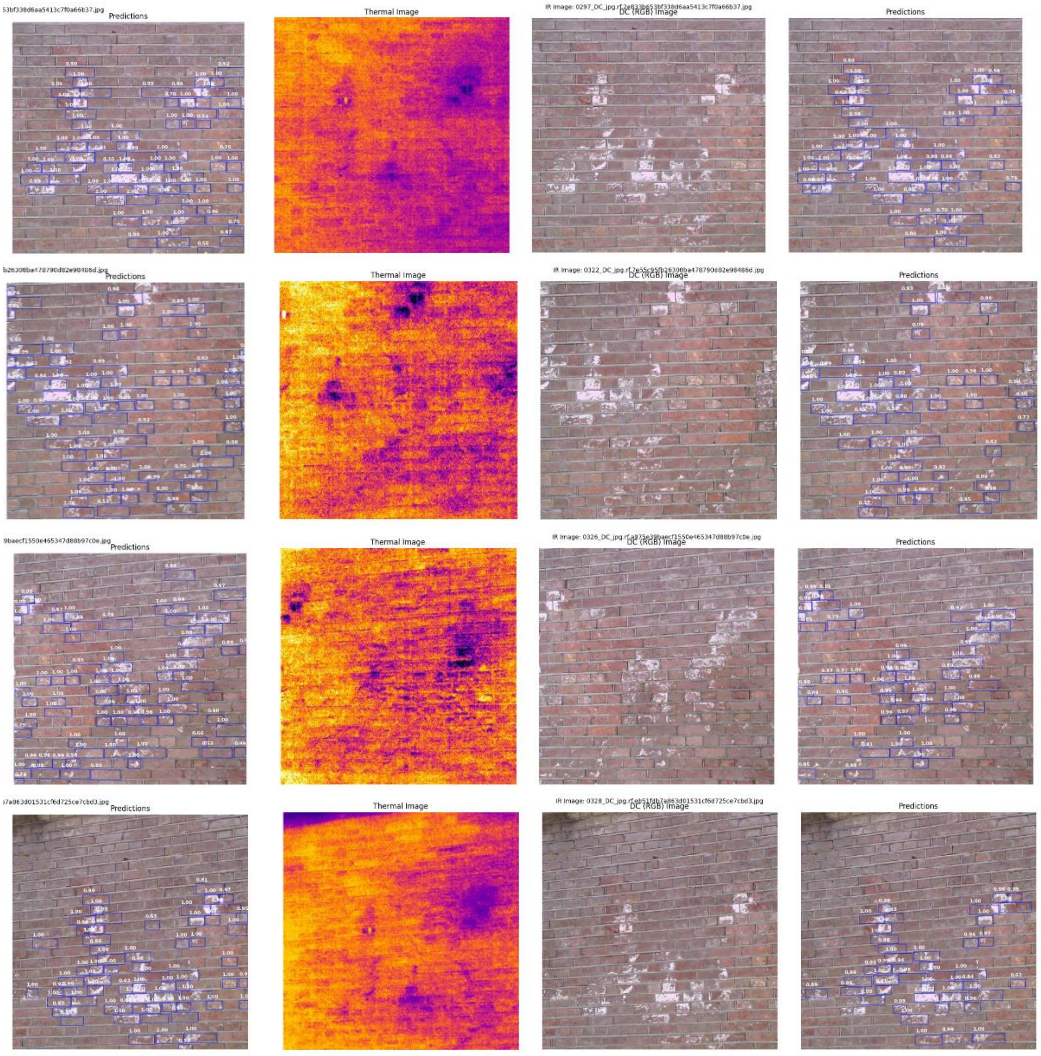
Tian, Y., Ye, Q., & Doermann, D. (2025). *YOLOv12: Attention-Centric Real-Time Object Detectors*. arXiv preprint arXiv:2502.12524.

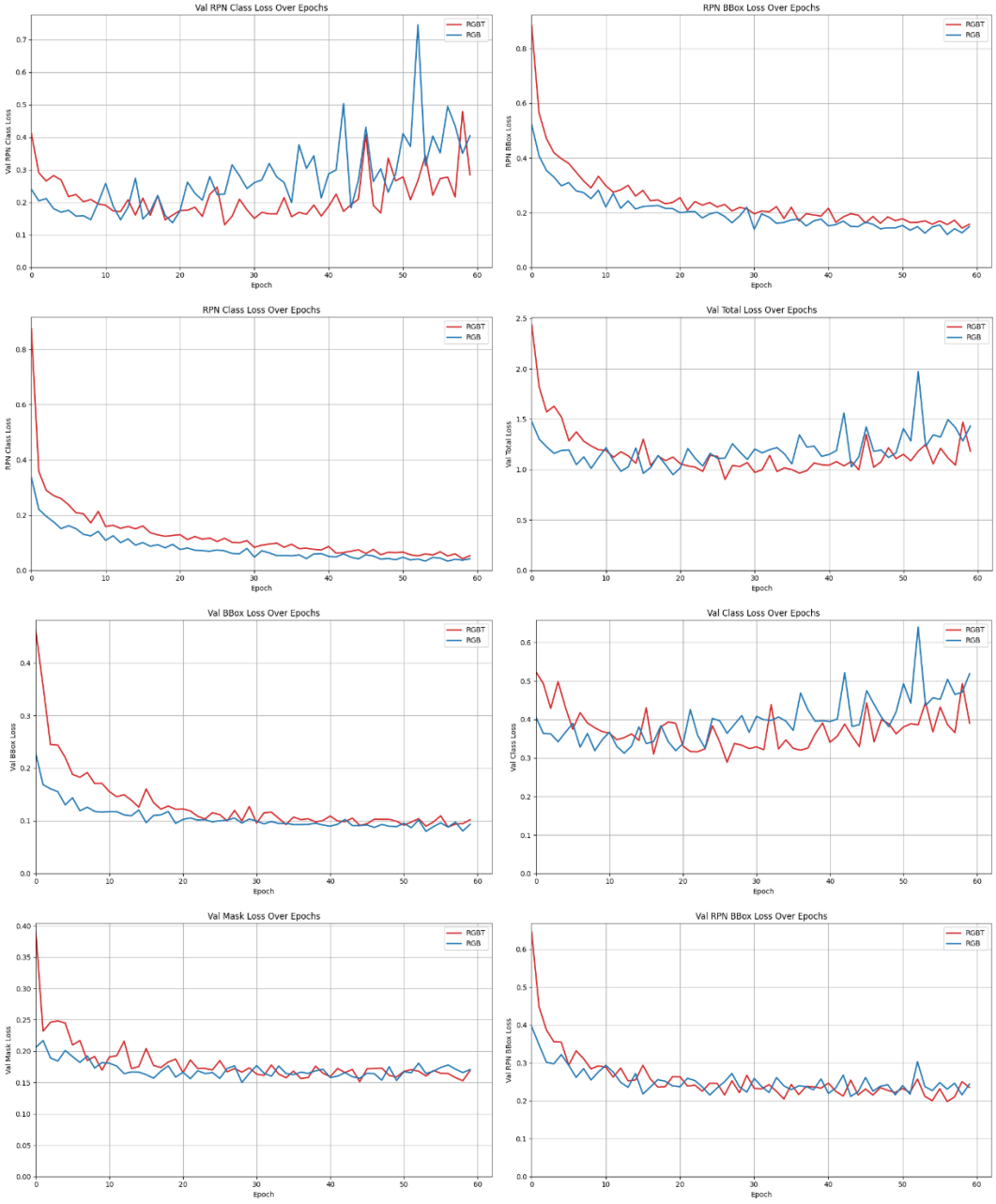
Appendices

Appendix I — Additional results

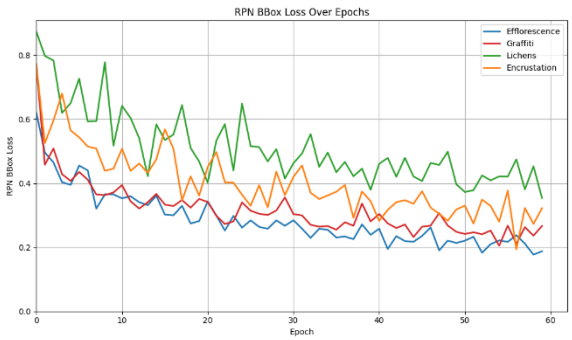
MODEL A - RGBT vs MODEL B – RGB

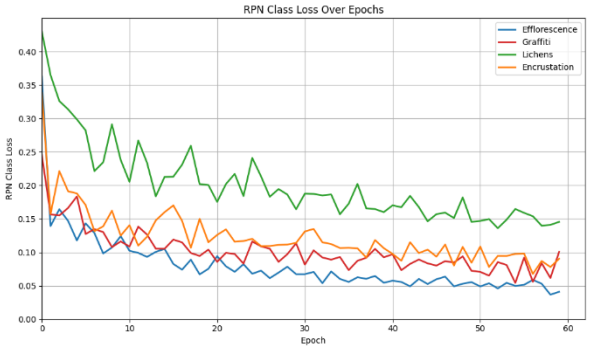
Upon request the entire thermal imagery dataset can be accessed

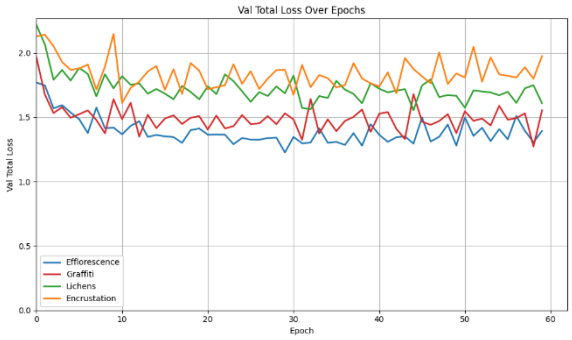


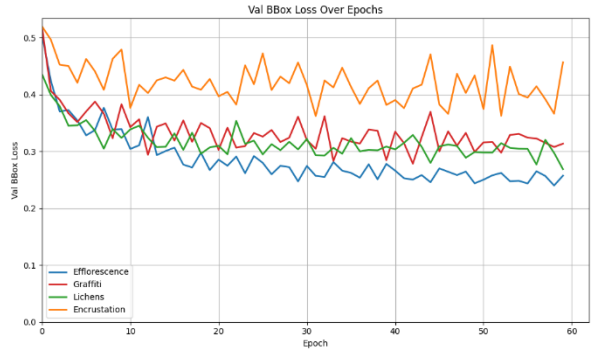

Val RPN Class Loss Over Epochs
RPN BBox Loss Over Epochs
RPN Class Loss Over Epochs
Val Total Loss Over Epochs
Val BBox Loss Over Epochs
Val Class Loss Over Epochs
Val Mask Loss Over Epochs
Val RPN BBox Loss Over Epochs
RGBT
RGB
Epoch

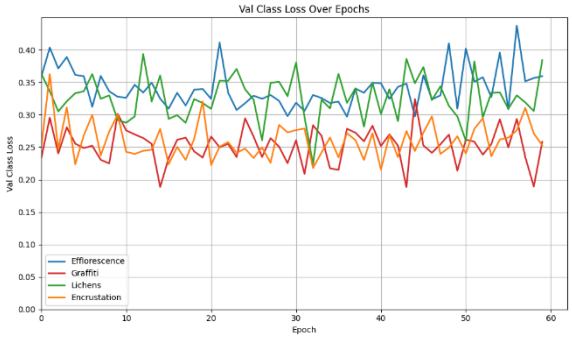
MODEL D - GRAFFITI vs MODEL E – LICHENS vs MODEL F – ENCRUSTATION

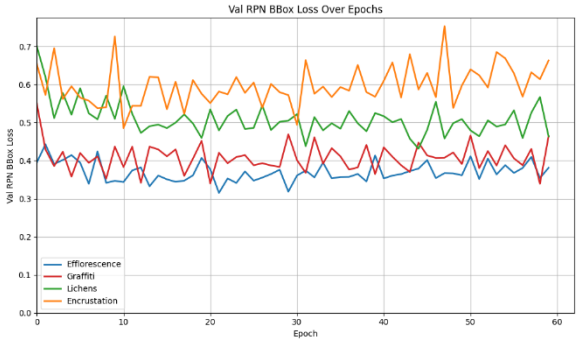


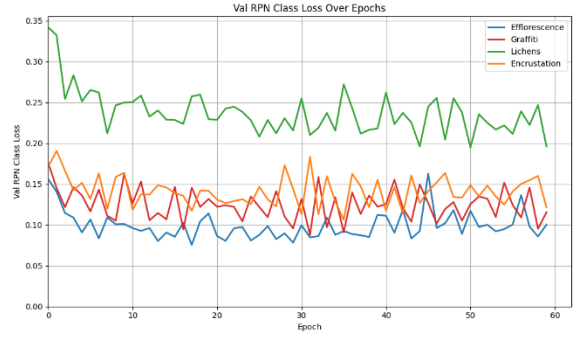




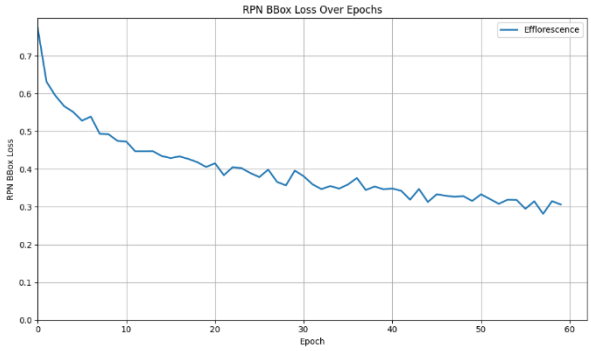


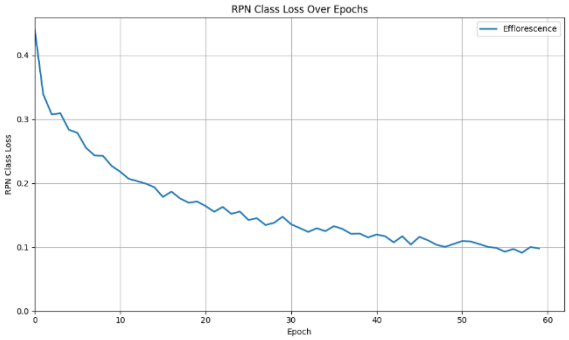


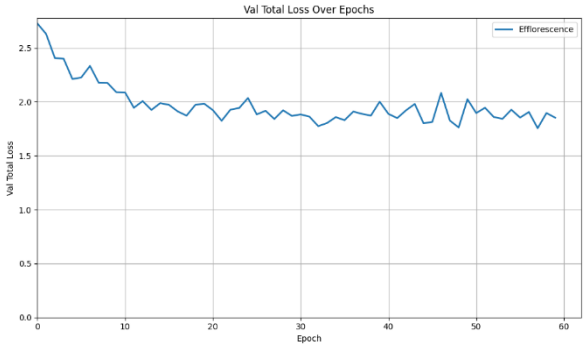


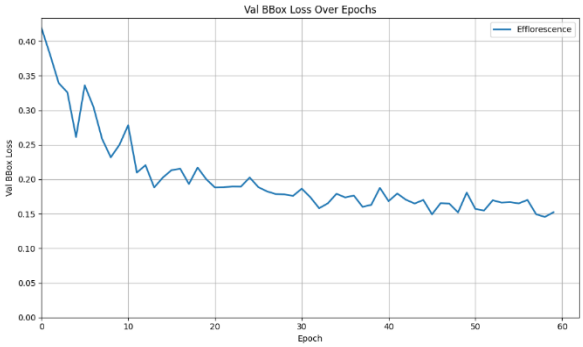


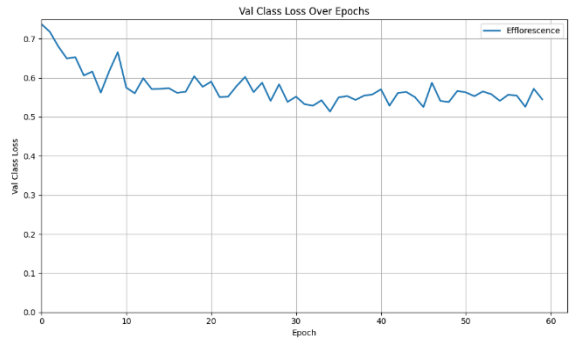
MODEL I – DAMAGE & EFFLORESCENCE

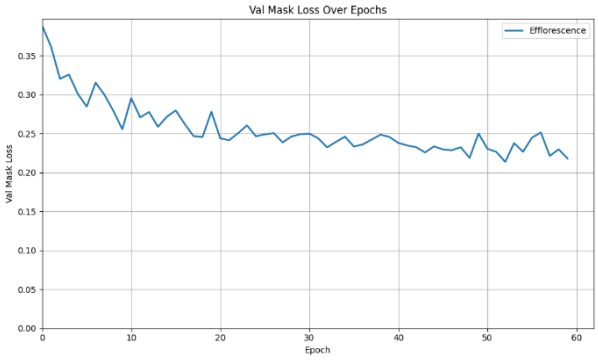


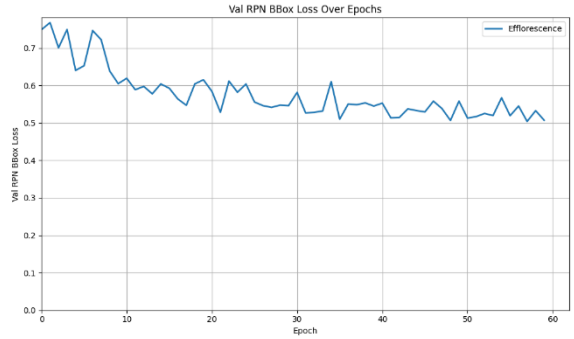


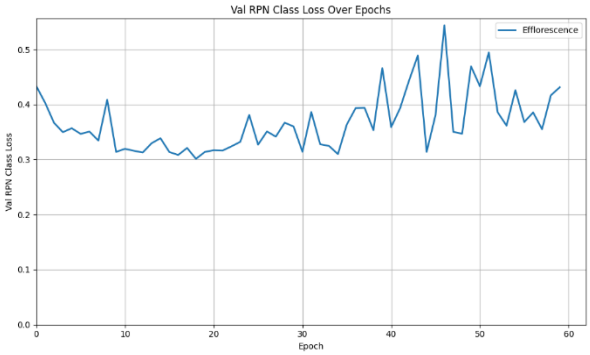












Chi-square per cumulative cutoff

Cutoff	In~Dam	In~Und	Out~Dam	Out~Und	Chi2	p-value	Sig<0.05
≤1.0	449	74	11748	4301	41.065	1.47E-10	YES
≤1.5	1239	297	10958	4078	43.0768	5.26E-11	YES
≤2.0	2061	553	10136	3822	43.6147	4.00E-11	YES
≤2.5	3513	1029	8684	3346	44.8911	2.08E-11	YES
≤3.0	4581	1448	7616	2927	27.497	1.57E-07	YES
≤3.5	6532	2151	5665	2224	24.6863	6.75E-07	YES
≤4.0	8002	2705	4195	1670	19.9321	8.02E-06	YES
≤4.5	10288	3642	1909	733	2.841	9.19E-02	no

Cummilatitve contingency table

Cutoff(≤×)	Eff~Damage	Eff~Undamaged	Clean~Damage	Clean~Undamaged	Row Total
≤1.0×	449	74	11748	4301	16572
≤1.5×	1239	297	10958	4078	16572
≤2.0×	2061	553	10136	3822	16572
≤2.5×	3513	1029	8684	3346	16572
≤3.0×	4581	1448	7616	2927	16572
≤3.5×	6532	2151	5665	2224	16572
≤4.0×	8002	2705	4195	1670	16572
≤4.5×	10288	3642	1909	733	16572

Contingency table

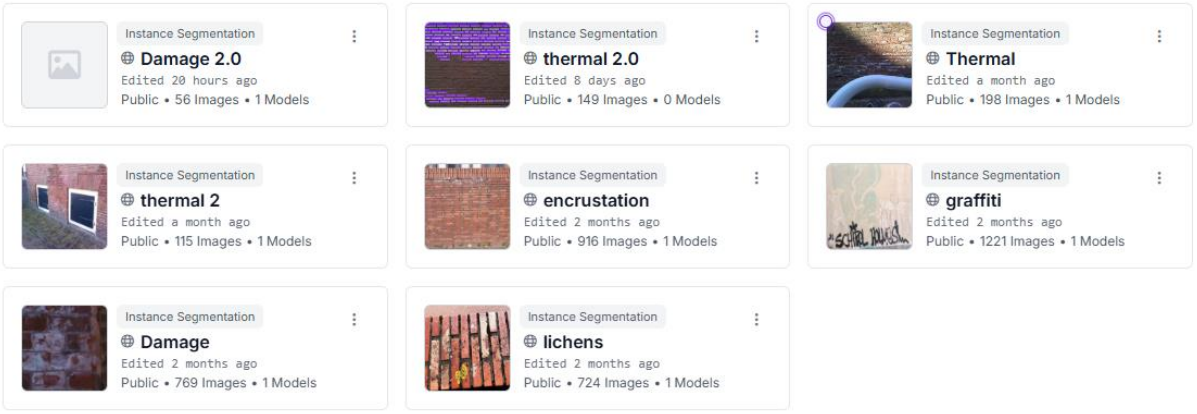
Zone	Damaged	Undamaged	Row Total (Ri)
Zone 1 (≤1.0×)	449	74	523
Zone 2 (1.0×–1.5×]	790	223	1013
Zone 3 (1.5×–2.0×]	822	256	1078
Zone 4 (2.0×–2.5×]	1452	476	1928
Zone 5 (2.5×–3.0×]	1068	419	1487
Zone 6 (3.0×–3.5×]	1951	703	2654
Zone 7 (3.5×–4.0×]	1470	554	2024
Zone 8 (4.0×–4.5×]	2286	937	3223
Zone 9 (4.5×–5.0×]	1909	733	2642
COLUMN TOTALS	**12197**	**4375**	**16572**

Appendix II – Script

All code is publicly available on GITHUB:

https://github.com/ValentijnCamielCloo/VCLOO_Master_Thesis

Appendix III — Dataset



All datasets are publicly available at

https://app.roboflow.com/valentijn/