

EXPLORING THE POTENTIAL OF DEEP LEARNING-BASED IMAGE ANALYSIS FOR DAMAGE RECOGNITION IN HERITAGE BUILDINGS

// REFLECTION

October 21st, 2025

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

Blank page

Colophon

TITLE	<i>Exploring the potential of deep learning based image analysis for damage recognition in heritage buildings</i>		
DOCUMENT VERSION	MSc. Graduation Thesis REFLECTION		
STATUS	1.0		
DATE	FINAL		
AUTEUR	OCTOBER 21 st 2025		
STUDENT NR.	V.C. (Valentijn) Cloo		
	5436958		
INSTITUTE	Delft University of Technology		
DEGREE	Master of Science Architecture, Urbanism and Building Sciences		
DEPARTMENT	Architectural Engineering + Technology		
TRACK	Building Technology		
ADRES	Julianalaan 134, 2628 BL Delft		
1 st MENTOR	Dr. Barbara Lubelli		
TITLE	Associate Professor		
CHAIR	Heritage & Architecture		
2 nd MENTOR	Dr.	Azarkhsh	Rafiee
TITLE	Assistant		Professor
CHAIR	Geographic Information System Technology		
EXAMINATION COMMITTEE	XXX		

Graduation Process

My graduation project focuses on the application of machine learning for the detection of efflorescence in heritage masonry buildings. This aligns directly with the studio's emphasis on integrating digital technologies within sustainable design strategies for the built environment. Positioned within the Building Technology track of the MSc Architecture, Urbanism, and Building Sciences, the project brings together material science, computational methods (Design Informatics), and conservation principles (Heritage & Architecture), highlighting the program's multidisciplinary nature. By contextualizing technical innovations within architectural heritage, the research aims to bridge scientific development and cultural preservation.

Research approach

The research methodology followed a four-stage approach, illustrated in Figure 1: (1) Literature Research, (2) Experimental Design, (3) Validation & Application, and (4) Reporting. Initially, I conducted an in-depth literature review on efflorescence and the wider spectrum of masonry damage types. Around week 4–5, the scope was refined using the MoSCoW method to define model requirements, leading to a structured outline of damage classes for detection. It was agreed early on that dataset quality and annotation consistency were critical to achieving reliable model performance.

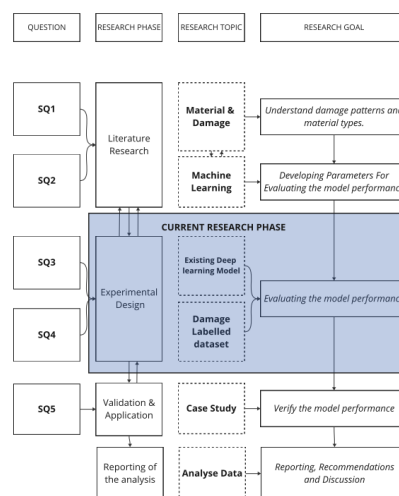


Figure 1 Research Methodology

During the data annotation phase, several important lessons emerged. I began by using general area annotations, but later moved to a brick-by-brick strategy to increase precision. While more detailed, this shift proved extremely time-consuming and led to unexpected complications, such as increased overfitting and imbalance in model training. I also experimented with labelling efflorescence in joints as a separate class, but this introduced significant class imbalance due to too few representative samples. This highlighted how annotation strategies can significantly influence model performance, and that precision alone isn't beneficial unless it is supported by balanced data and sufficient sample size.

Dataset and annotations

Efforts were made to reach out to scholars internationally (e.g., China, Spain, Turkey) to gather datasets and insights, but responses were limited. The initial dataset shared by the heritage supervisor contained fewer than 100 images, and given the known constraints of machine learning performance with small datasets, it became evident that I needed to collect additional data through fieldwork. This broadened the project scope and required me to understand real-world detection practices. Unfortunately, from the 12 Dutch contractors approached, only 3 responded. Those who participated lacked structured methodologies for efflorescence detection, reinforcing the need for innovation in this field.

Simultaneously, I explored state-of-the-art machine learning methods and performance evaluation metrics. This dual-track literature review—focusing first on damage phenomena, then on AI techniques—allowed me to identify research gaps and define the novelty of my approach. However, one methodological weakness was the limited initial focus on annotation strategies, which had to be revisited once poor model performance (overfitting, underfitting, class imbalance) became evident. Furthermore, I found that academic sources often lacked detail on practical setbacks, particularly regarding image scale and field conditions—critical issues that were frequently oversimplified.

From a technical standpoint, the project encountered several barriers. Frequent coding issues, GPU memory crashes, and inconsistencies in training—especially with high-resolution data and multi-class setups—often disrupted workflow. In particular, Mask R-CNN models struggled with mask loss not being properly trained, and the introduction of additional classes added too much complexity relative to the dataset size and project timeframe. These issues taught me the value of setting a realistic scope and embracing iteration as a core part of applied machine learning research.

Early training results showed strong overfitting, which led me back to the literature. There, I discovered that many successful studies used tightly controlled image capture conditions e.g., consistent 60 cm distance, orthogonal camera angles, uniform lighting, and large datasets. In my case, capturing field images under such standardized conditions was not feasible. This raised the issue of scale sensitivity in real-world damage detection something rarely emphasized in academic papers. A potential future solution involves implementing a sliding window or tiling method to normalize image scale and context during model training.

These setbacks shifted my perspective from model-centric evaluation to a more critical, data-focused approach. I learned to evaluate the quality of input data and annotation strategies as foundational pillars for success. Diving deeper into external studies helped clarify why overfitting occurred, and revealed the environmental and technical constraints needed for deep learning models to work reliably in real-world settings.

During the model training process, I encountered a significant issue that impacted the validity of my multi-class detection results. My goal was to develop a model capable of detecting both efflorescence and other classes, building on the improved accuracy achieved for efflorescence detection, which increased from 0.60 to 0.85 on the YOLOv8 model. To achieve this, I fine-tuned a new model for multi class detection using the pre-trained weights from my previous efflorescence model. However, I did not take into account that the model architecture and class

mappings were initially configured for detecting single class only. As a result, during training, the model effectively replaced the efflorescence class with the newly developed class, rather than learning to detect both types of damage. This mistake occurred because I did not properly configure the model to accommodate both classes, and instead, the model retained the single-class structure from the efflorescence training.

The consequence of this error is that the newly trained model no longer recognizes efflorescence, thereby undermining my ability to perform accurate multi-class validation and evaluation. This issue directly affects the primary objective of my research, which focuses on improving multi-class damage detection and mitigating misclassification risks.

To resolve this, I will retrain the model using the COCO pre-trained weights instead of the efflorescence weights, ensuring that the model correctly learns multiple classes. Although this mistake delayed the completion of the intended results for the concept report, I am committed to resolving the issue promptly and ensuring that the final model is robust and reliable for both damage types.

Research and Design Integration

The relationship between research and design in this project is cyclical. The model design (e.g., architecture, parameter tuning) was driven by insights from literature and performance outcomes. For instance, after testing an initial model with single-class (efflorescence-only) annotations, I introduced multi-class labels and adjusted the annotation strategy (e.g., brick-by-brick annotation) based on performance analysis. These changes significantly impacted model behavior. Mask R-CNN improved with the refined annotations, while YOLOv8 exhibited performance loss due to class imbalance.

Experiments also incorporated various parameters identified through literature—image resolution, environmental lighting, viewing angles, spatial context, and near-infrared imagery—to assess their impact on model accuracy. Each adjustment to the model was directly informed by research findings, creating a continuous feedback loop between experimentation and theory.

Throughout this process, I developed hands-on skills in Python debugging, dataset formatting (e.g., COCO format), annotation tools (like LabelMe and Roboflow), and model configuration. These technical abilities were built gradually through trial and error, further emphasizing the importance of flexibility and continuous learning when working with complex, evolving tools like Mask R-CNN.

Throughout the development of my thesis, I encountered several challenges that required additional research and reflection, integrating both the technical and research aspects of the project. One of the key moments where I had to delve deeper into research was when I realized the impact of the time of day on thermal imagery. Initially, I did not account for the significant variation in thermal readings caused by changing environmental conditions, particularly sunlight exposure. This prompted me to investigate optimal measurement parameters to ensure consistency and accuracy in data collection. As a result, I gained a better

understanding of how to calibrate the model to account for these variations, which significantly improved the reliability of the thermal data.

Another challenge emerged when addressing the scaling issue within the dataset. Due to the limited and inconsistent data provided from both online sources and my supervisor, I had to shift my focus toward generating a more diverse and representative dataset. This required extensive additional research into the types, causes, and patterns of efflorescence damage. While this was initially a hurdle, it ultimately enriched my understanding of the damage mechanisms and equipped me with the knowledge to assess damage accurately during site visits. This not only improved my practical insight but also allowed me to engage more meaningfully with stakeholders and provide informed opinions on preservation strategies.

During the integration of thermal and DC camera images, I encountered unexpected challenges related to the alignment and calibration of the image pairs. This required me to research thermal imaging settings, particularly the importance of distance calibration and sensor alignment to ensure accurate data fusion. Moreover, distinguishing between rising damp and leakage using thermal data posed an additional layer of complexity. Since rising damp typically manifests as cooler areas near the ground, while leakage often appears as cooler spots higher on the wall, training the model to differentiate these conditions required careful annotation and model adaptation.

Finally, balancing training efficiency with model accuracy proved to be an ongoing consideration. As memory limitations frequently led to model crashes, I had to make critical decisions regarding image resolution, batch size, and training duration. This iterative process taught me the importance of balancing computational demands with the need for high-quality input, reinforcing the interconnected nature of technical constraints and research outcomes.

In summary, the integration of research and design throughout the project was not a straightforward process but rather an evolving journey that required constant reflection and adaptation. These challenges not only enhanced my technical skills but also deepened my understanding of the complex factors influencing efflorescence detection, ultimately shaping the quality and reliability of my model.

Ethical and Moral Considerations

The main ethical dilemma encountered related to dataset ownership. During discussions with other researchers and institutions (e.g., TNO, TU Delft), it became apparent that reusing existing datasets was complicated by unclear image rights and licensing issues. As a result, I focused on collecting my own field data, supplemented by additional datasets provided through my supervisors' networks. This approach ensured that data usage remained transparent and ethically sound while also enhancing dataset diversity.

To build a robust and diverse dataset, I conducted field visits to various heritage buildings, including churches and other historic structures (bridges), in different cities across the Netherlands. Some of the key locations visited include:

- Amsterdam: Old Church, Hodson Dedeul Square

- Katwijk: Andreas Church
- The Hague: Orange Hotel, Scheveningen National Archives
- Wateringen: Square Church
- Leiden: Waalse Church
- Delft: Jesse Maria Church

By collecting data from these diverse sites, I aimed to capture a wide range of efflorescence patterns and related damage types under varying environmental and architectural conditions. This proactive data collection effort not only addressed ethical concerns but also strengthened the quality and generalizability of the dataset used for model training and validation.

Societal Impact

Practical Applicability

The practical applicability of the research lies in its potential to automate and standardize masonry damage assessments. For example, a drone could capture systematic imagery of the masonry tower at the Faculty of Architecture. These images could then be processed by the trained model to identify defects, as defined by the MoSCoW damage classification. This application demonstrates how machine learning can facilitate non-invasive, scalable inspections in both academic and professional contexts.

Innovation Achievement

Several innovations were realized during this project. The use of both YOLOv8 and Mask R-CNN allowed for object detection and instance segmentation, respectively. While Mask R-CNN is a well-established architecture, its application to efflorescence detection, especially under variable field conditions, has not been extensively studied. This includes scenarios where images are captured at non-orthogonal angles or without scaffolding, such as inspecting the roof of a historic church. My research also introduces a novel integration of moisture-related parameters using thermal imaging. Unlike most prior studies that assume a dry surface or use laboratory settings, this project bridges thermal data and efflorescence detection, proposing an interdisciplinary methodology for field-based assessment.

Contribution to Sustainable Development

The project contributes to sustainable development in several ways. Socially, it supports the conservation of heritage assets and cultural identity by providing early detection methods to prevent structural deterioration. Environmentally, it

promotes preservation over demolition, reducing the material and energy costs associated with full renovations. Economically, automating inspections could lower operational costs and enhance maintenance planning, which aligns with sustainable business practices in architecture and heritage management.

Impact on People, Planet, and Prosperity

- **People:** Supports the preservation of culturally significant structures, enhancing community identity and engagement.
- **Planet:** Reduces the environmental impact of unnecessary renovations or demolitions through early detection and targeted interventions.
- **Prosperity:** Offers cost-effective diagnostic tools, potentially reducing labor-intensive inspections and enabling preventive maintenance.

Socio-Cultural and Ethical Impact

By enabling timely and accurate damage detection, the research indirectly supports the longevity of historically and culturally valuable buildings. It also addresses an ethical gap in current practice—many contractors lack defined strategies, and decisions are often based on subjective judgment. This project introduces a more objective and data-driven approach, contributing to professional ethics in heritage care.

During my visits to multiple churches across the Netherlands, I observed that many of these structures are in a deteriorated state, often with limited financial support that predominantly focuses on maintaining the exterior facades. In some cases, I engaged with local communities and church representatives, who shared insights into the challenges they face. In cities like Amsterdam, larger churches have adapted to financial pressures by transforming into museum-like spaces, attracting tourists and generating income.

Furthermore, I observed that the state of church maintenance and monitoring remains limited. While crack sensors were often present, it became apparent that these have not been maintained for years, reflecting a lack of consistent building condition monitoring. Additionally, conversations with building managers revealed that their architectural knowledge is often limited, and they frequently struggle with insufficient time and resources to maintain the churches. One notable example shared with me was the cleaning and maintenance of ornate windows and decorations, where each ornament required about a week of dedicated work, posing a significant challenge given the limited manpower.

Some solutions applied to manage damage were surprisingly simplistic and not always effective. For instance, in cases of cracking and paint flaking due to salt deposition, the temporary solution involved covering the affected areas with curtains and cloths rather than addressing the underlying cause. Additionally, local residents mentioned that groundwater levels and the proximity of industrial activities might contribute to damage, but building managers, due to a lack of technical expertise, tend to respond only reactively to visible risks rather than implementing sustainable solutions. These experiences during my research visits have deepened my commitment to the preservation of monumental buildings, and

I intend to maintain contact with the communities to support their ongoing maintenance efforts.

Relation to the Wider Social Context

Efflorescence and moisture-related damage are not only technical issues but also social ones, affecting the usability and perception of public spaces. Improving diagnostics helps prioritize interventions, supporting safer and more dignified environments for communities that live with or near historic architecture. During the site visits I noticed a stark contrast in smaller, more conservative church communities where the lack of visitor engagement and local support leads to financial struggles. For example, during my visit to Delft, I learned that two out of the seven churches in the city are facing potential sale due to financial difficulties. Ironically, despite local petitions advocating for the preservation of these churches as community spaces, the lack of regular visitors continues to jeopardize their future.

Impact on Architecture and the Built Environment

The project offers a new digital toolset for architects, conservators, and engineers engaged in the preservation of the built environment. By merging digital diagnostics with material understanding, it enhances decision-making in design, renovation, and conservation practices. It exemplifies how technological innovation can reinforce rather than replace architectural expertise and help where there are limited financial resources to maintain heritage masonry.