buildings

MDPI

Article

Artificial Intelligence for Routine Heritage Monitoring and Sustainable Planning of the Conservation of Historic Districts: A Case Study on Fujian Earthen Houses (Tulou)

Jiayue Fan [1], Yile Chen [2,*] and Liang Zheng [2,*]

[1] Faculty of Innovation and Design, City University of Macau, Avenida Padre Tomás Pereira, Taipa, Macau 999078, China; u23092120461@cityu.mo
[2] Faculty of Humanities and Arts, Macau University of Science and Technology, Tapai, Macau 999078, China
* Correspondence: 2009853gat30001@student.must.edu.mo (Y.C.); 2009853gat30002@student.must.edu.mo (L.Z.)

Abstract: With its advancements in relation to computer science, artificial intelligence has great potential for protecting and researching the world heritage Fujian earthen houses (Tulou) historical district. Wood is an important material used in the construction of Fujian earthen houses (Tulou); wood is used in both the main structure of the buildings and for decoration. However, professionals must invest significant time and energy in evaluating any damage before repairing a building. In this context, this study proposes and optimizes a detection method based on the YOLOv8 model for detecting damage to the wooden structure of Fujian earthen houses. Through multiple experiments and adjustments, we gradually improved the detection performance of the model and verified its effectiveness and reliability in practical applications. The main results of this study are as follows: (1) This machine-learning-based object detection method can efficiently and accurately identify damaged contents, overcoming the limitations of traditional evaluation methods in terms of labor and time costs. This approach will aid in the daily protection monitoring of historical districts and serves as a preliminary method for their renewal and restoration. (2) Through multiple rounds of experiments, we optimized the YOLOv8 model and significantly improved its detection accuracy and stability by removing samples with complex backgrounds, improving label quality, and adjusting hyperparameters. In the final experiment, the model's overall mAP was only 57.00% at most. However, during the field test, the model successfully identified nearly all damage points, including holes, stains, and cracks in the wooden structure of the analyzed earthen building, effectively fulfilling the requirements of the detection task. (3) In the KuiJu Lou field test in Fujian Tulou, the model also performed well in complex environments and was able to reliably detect damage types such as holes, stains, and cracks in the wooden structure. This test confirmed the model's efficiency and stability in practical applications and provided reliable technical support for Fujian Tulou protection and restoration.

Keywords: machine learning; world heritage; YOLOv8; historic district protection; earthen houses (tulou); Fujian

check for updates

Citation: Fan, J.; Chen, Y.; Zheng, L. Artificial Intelligence for Routine Heritage Monitoring and Sustainable Planning of the Conservation of Historic Districts: A Case Study on Fujian Earthen Houses (Tulou). *Buildings* **2024**, *14*, 1915. https://doi.org/10.3390/buildings14071915

Academic Editor: Hugo Rodrigues

Received: 30 May 2024
Revised: 19 June 2024
Accepted: 20 June 2024
Published: 22 June 2024

Copyright: © 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by/4.0/).

1. Introduction

1.1. Research Background

The architectural structures known as Fujian Tulou (earthen houses) are primarily situated within the Fujian Province, China. In a wider sense, Tulou architecture is dispersed across the southeast and south of China, with a smaller quantity distributed in Southeast Asia. UNESCO has recognized 46 Tulous in the Fujian Province as "outstanding examples of traditional architecture and function" and listed them as World Heritage Sites [1]. Tulous, with a history spanning several dynasties and nearly 800 years, ranging from the Yuan Dynasty (A.D.1271–A.D.1368) through to the latter half of the 20th century [2], have

had their main characteristics shaped and defined by local Hakka and Minnan residents, who have adapted their construction to the local conditions [3]. The emergence of the Tulou is associated with the turbulent political climates of the late periods of various dynasties. The origin of Tulou architecture in Fujian dates to the late Song Dynasty (A.D.960–A.D.1279) and the early Yuan Dynasty (A.D.1271–A.D.1368), with a significant increase in Tulou construction towards the end of the Ming Dynasty (A.D.1488–A.D.1605) [4]. These Tulous represent a unique form of communal living and defensive organization in the mountainous regions of Southern China. Moreover, their harmonious interaction with the environment positions them as exemplary models of human habitation [5].

In the past, China extensively used architectural structures of a similar form, known as Wubao, as military defensive installations [6]. This type of architecture reached its zenith during the Wei, Jin, Southern, and Northern Dynasties (3rd to 6th centuries AD) [7]. Wubao construction often featured watchtower layouts, and the choice of building materials differed from that of Tulous [8]. This aspect contrasts with the more uniform architectural layout and material usage in Tulous. With the establishment of the Sui, Tang, and other dynasties after the 6th century, the prevalence of Wubao-style architecture in Northern China gradually waned due to its militaristic nature, making it a primary target for government repression [9]. As warfare in the northern regions led to the southward migration of the Hakka people, these people combined the architectural techniques of the Wubao with local materials, marking a new phase of development for the Tulou. The earliest written record of Tulous dates to the Yuan Dynasty (A.D.1271–A.D.1368) [10]. According to the existing literature, the Tulou of this period closely resembled the architectural concepts and techniques of the Wubao. During the Ming and Qing Dynasties (A.D.1368–A.D.1912), Tulou architecture further evolved, introducing a round form and significantly increasing in scale. The transition from simple rammed-earth materials to a combination of earth and wood, as well as a mixture of earth and stone, marked the beginning of a more complex construction process. The extensive use of wooden materials shifted the function of these structures from military to residential, integrating clan and community living with cultural significance [11]. As a significant form of architectural cultural heritage, the heavy reliance on wooden materials renders Tulous in Southern China particularly susceptible to damage from natural causes [12]. Additionally, the daily accommodation of large populations and the influx of tourists have led to an increase in damage to the wooden structures. If not detected and promptly repaired, this damage could potentially lead to catastrophic safety hazards. With the development of big data and artificial intelligence technology, it has become possible to conduct non-destructive testing and regular heritage monitoring in historical districts such as Fujian Tulou.

1.2. Literature Review

1.2.1. Development of Historical District Protection

Historic districts serve as a testament to urban development, embodying a wealth of cultural heritage and historical memory [13,14]. The preservation and monitoring of these districts are not only expressions of respect for historical culture but also critical components of sustainable urban development [15,16]. Early preservation efforts in historic districts primarily focused on the physical form of the buildings themselves, emphasizing the repair and restoration of historical buildings to revive their original appearance [17]. Early preservation was characterized by several key features. There was an emphasis on authenticity: the pursuit of restoring a building to its initial state, utilizing original materials and techniques as much as possible [18]. The main focus of the protection efforts was on standalone buildings of significant historical value, while the overall urban historical architecture received less attention [19]. While this approach to preservation protected the physical form of historical buildings to some extent, it also had its inherent limitations.

With societal progression and the evolution of preservation philosophies, the modern preservation of historic districts places greater emphasis on sustainable development [20],

accentuating the balance between protecting cultural heritage and considering social, economic, and environmental benefits [21].

Since the concept of protecting buildings and areas of historical value was first proposed in the *1933 Athens Charter*, the philosophy surrounding the preservation of historical districts has progressively evolved and deepened. By the time of the *1976 Nairobi Recommendation*, historical districts were depicted as vivid representations of diverse cultural, religious, and social characteristics, emphasizing the importance of preserving these areas for future generations [22]. This recommendation also highlighted that integrating the preservation of historical districts with modern social life should become a core component of urban planning and land development.

Further developments were made in 2005 with the *Vienna Memorandum and the Declaration on the Conservation of Historic Urban Landscapes*, which expanded the concept of historic landscape conservation. These documents moved beyond the traditional confines of historic centers and building complexes to include a broader range of land and landscape environments, marking a significant expansion of the preservation philosophy for historical areas. Collectively, these documents underscore the pivotal role of historical districts in maintaining cultural heritage and fostering social identity while also reflecting the evolution of preservation strategies to meet the demands of modern urban development.

The use of architectural heritage as an economic asset to attract investment is currently trending [23], posing challenges in terms of determining how best to protect cultural heritage within a limited budget. Moreover, the influx of a large number of visitors can lead to building damage; hence, timely and real-time assessment for the protection of an entire district has become a complex issue. Simultaneously, modern cultural heritage protection has evolved from protecting individual entities to safeguarding entire districts or regions [24]. In recent years, with the advancement of computer technology, scholars have begun to apply machine learning and other technical means to the preservation of historic districts. This approach has further allowed for the efficient maintenance of buildings and other forms of cultural heritage within these districts [25–30]. In the past, the costs associated with the maintenance inspection and subsequent preservation of related cultural heritage buildings were substantial [31,32]. The application of technologies such as YOLO (You Only Look Once) in machine learning can further reduce the labor costs of building inspections and related maintenance [33,34], thereby lessening the conservation burden placed on large-scale building communities.

1.2.2. The Application of Computer Vision Technology to Architectural Heritage

Computer vision is an artificial intelligence technology that endows computer systems with human-like image recognition capabilities [35]. In recent years, the YOLO (You Only Look Once) series of algorithms has stood out in the field of computer vision due to its real-time performance and high accuracy [36]. The YOLO algorithm transforms the object detection problem into a regression problem, achieving end-to-end object detection through a single convolutional neural network [37]. Subsequently, researchers have introduced versions such as YOLOv2, YOLOv3, YOLOv4, and YOLOv5 [38], which further enhance the performance and applicability of this algorithm. The YOLO series has been widely applied in areas such as security monitoring [39], autonomous driving [40], medical imaging [41], and industrial quality inspection [42]. Looking forward, the development of the YOLO algorithm may include integration with other technologies and expansion into more application fields [43].

The application of computer vision technology in the field of architectural and cultural heritage is a rapidly evolving research area. This technology, which uses image-processing and analysis algorithms, can help us better understand, protect, and manage these invaluable cultural assets. For instance, K. Siountri and colleagues employed a deep learning-based image classification method to classify and recognize historical buildings in Athens, achieving favorable results [44]. Perez-Toro and others proposed a convolutional-neural-network-based method for detecting damage to buildings, providing important references

for building maintenance [45]. Some scholars have used computer vision technology to investigate archaeological sites [46]. M. Altaweel and colleagues utilized high-resolution images captured by drones, combined with semantic segmentation algorithms, to effectively identify different ground features in archaeological studies [47]. Jalandoni, A. and others developed a deep learning-based method for detecting rock art, greatly enhancing the efficiency of archaeological research [48].

The current applications of the YOLO algorithm across various fields, such as security monitoring, autonomous driving, and architectural heritage preservation, exhibit consistent similarities. Notably, they replace traditional manual recognition and operations with a more advanced and economical approach, addressing potential issues inherent in these traditional methods [49,50]. Furthermore, the adoption of deep learning and related techniques enhances these methods' long-term scalability and future upgrade capacity. Computer vision technology is strengthening and becoming widely applied, offering new tools and methods for research, protection, and management in the fields of architectural and cultural heritage preservation. With the ongoing development of this technology, researchers can anticipate innovative applications of computer vision in more fields, making a significant contribution to the preservation and development of cultural heritage.

1.2.3. The Use of Deep Learning in the Sustainable Preservation of Wooden Structures

Wood is a traditional building material which may retain a sound appearance over time, yet it is subject to risks of biological degradation or excessive internal wear and tear [51]. These potential issues usually only come to light during architectural disasters or structural replacements, highlighting the importance of wood durability and structural integrity in ensuring architectural safety. For instance, end rot can cause the beams supporting a ceiling to lose their wall support, a situation that is difficult to detect without periodic and thorough inspections [52].

The use of YOLO (You Only Look Once) visual recognition technology can significantly improve monitoring and diagnostic efficiency for wooden structures [53]. YOLO is an efficient object detection algorithm that can quickly identify and classify various objects in real-time video streams [54]. We can effectively apply this technology to architectural areas that are difficult to inspect directly, such as roof beams and floor support structures, by training the YOLO algorithm to detect signs of damage in wood, such as cracks, decay, and warping. The implementation of YOLO technology not only helps increase detection speed but also enhances detection accuracy [55]. Early wood defect detection mainly relied on stress wave, X-ray, and ultrasonic technology. Stress wave technology requires high environmental control, and the equipment used needs to be fixed on the wood. X-ray imaging limits the sensitivity of defect identification due to its poor image quality and low visual contrast. Lastly, although ultrasonic waves can reveal internal and external defects, they are sensitive to external conditions and easily interfered with. These methods have their own advantages and disadvantages, which affect the accuracy and efficiency of detection [56]. Therefore, visual judgment is widely used as a standard for early detection and damage judgment for wood materials, and these preliminary assessments provide an effective method professionals can use to conduct targeted inspections later with minimal impact on buildings. YOLO technology can be deployed faster and provide preliminary visual judgments for early wood defect detection [57,58].

This technology enables building maintenance personnel to quickly detect hidden structural problems, facilitating early intervention and reducing potential safety risks. Moreover, YOLO's real-time processing and high recognition rate make it highly applicable in the field of architectural health monitoring [59], especially at a time when the demand for the continuous monitoring of architectural structures is increasing. Implementing YOLO technology in a real-time monitoring system can significantly reduce reliance on traditional manual inspections [60], and its early-warning system can automatically send notifications to the maintenance team upon the identification of potential risks. Its early

diagnostic features and capacity for timely interventions dramatically enhance building safety and operational efficiency [61].

Comprehensive and meticulous inspections, especially of the most vulnerable parts of wooden structures, are more critical than simply replacing damaged parts. Such detailed detection not only helps preserve the historical value of buildings, but, more importantly, ensures the safety of buildings and the structures in use [62,63]. Accurate diagnosis ensures safe structural operation, minimizing the need to reinforce, replace, or repair damaged parts and their adjacent elements. Coupling this diagnosis with YOLO visual recognition technology allows the construction industry to monitor the state of wooden structures accurately and in real time. The application of this technology not only helps reduce costs but also plays a significant role in environmental protection and sustainability [64]. Adopting this type of monitoring method underscores the importance of preventive maintenance. Compared to replacement or demolition, it can significantly reduce the demand for new raw materials and decrease energy consumption in the transportation, production, installation, maintenance, and recycling of buildings or their components, thus supporting sustainable development goals.

In addition to YOLO vision recognition technology, deep learning has been extensively applied in the field of building health monitoring. Strain in engineering structures describes the change in an object's unit length following a force application. It is a dimensionless quantity that represents a relative measurement of a material's degree of deformation. Strain usually manifests itself as the stretching or compression of an object in a local area, reflecting the stress state of the material's internal structure. Deformation refers to the displacement and shape change in an object as a whole in space. The change in position between the parts of the object enables the observation and measurement of its macroscopic manifestation. Usually, deformation corresponds to structural changes like bending, twisting, or compression. For instance, in civil engineering, researchers have utilized convolutional neural networks (CNNs) to automatically identify and classify cracks on concrete structural surfaces, significantly enhancing the efficiency and accuracy of crack detection. In the domain of steel structural bridges, scholars have proposed a method based on Bidirectional Long Short-Term Memory (BiLSTM) networks. This method can be used to learn the temperature–strain distribution characteristics of steel box girders under non-uniform temperature fields, providing an in-depth perception of structural strain field redistribution [65]. While YOLOv8 itself focuses on identifying and localizing objects in images, it can be trained to identify cracks or other signs of damage in the wooden structure of the earthen building, thereby indirectly helping to monitor strain and deformation. However, directly detecting strain and deformation usually requires combining sensor data or specialized image processing techniques rather than visual object detection models alone [66,67]. These research outcomes demonstrate that deep learning techniques can effectively process massive quantities of monitoring data and uncover the intrinsic patterns of structural responses, serving as a crucial supplement to traditional manual inspection methods.

Given that wooden structures also face issues such as aging degradation and biological decay, deep learning holds promising prospects for the health assessment of wooden constructions. By deploying sensors to measure temperature, humidity, and strain at critical points of wooden structures, and utilizing deep learning algorithms to analyze long-term monitoring data, it is feasible to detect internal anomalies in wood in a timely fashion, thereby providing early warnings of potential structural safety hazards. This approach can overcome the limitations of visual inspections and provide a macroscopic evaluation of the state of wooden constructions. If integrated with localized detection technologies like YOLO, it could form a comprehensive, multi-scale health-monitoring system for wooden structures, offering significant technical support for the sustainable protection of wooden constructions. However, due to the significant differences in the material properties of wood and steel, the application of deep learning in the wooden structure realm requires

extensive exploration and optimization, presenting a broad scope for innovation to researchers in the related fields.

1.3. Problem Statement and Objectives

According to field survey data, wooden components in Fujian Tulous generally suffer from varying degrees of damage. During the repair process, maintenance personnel frequently need to strike a balance between preserving original components and replacing damaged ones in order to preserve a Tulou's historical authenticity to the maximum extent. In light of the generally limited maintenance funds available for Tulou and the current context wherein the upkeep and operation of Tulous within tourist areas are managed by local cultural and tourism companies, the financial constraints have necessitated multifaceted considerations in the corresponding preservation efforts. As Tulous have developed over recent years, there is a compelling need to extend the lifespan of their wooden components that are not severely damaged as much as possible. At the same time, the adherence of newly replaced wooden components to standard specifications under a low-budget maintenance regime is subject to scrutiny. This situation underscores the complex decisions involved in preserving these historical structures while navigating the challenges of limited funding and regulatory compliance.

This dilemma poses larger challenges for Tulou protection and repair work. Over time, inadequate protection of wooden components may lead to structural stability issues or even partial collapses. To enhance the intelligent management and sustainable development ability of Tulou construction, in this study, we attempted to introduce machine learning methods to achieve the rapid detection and real-time monitoring of damage to wooden components. By summarizing the types of damage to Tulous' wooden components and establishing a corresponding image database of this damage, training intelligent recognition models can assist managers in promptly discovering and locating problems with wooden components, thereby allowing the formulation of targeted maintenance and repair plans. We expect that this intelligent monitoring method will enhance the efficiency and accuracy of Tulou protection, offering robust technical support for the sustainable use of Tulous. At the same time, we can also apply this method to other historical building types, offering fresh concepts and resources for safeguarding cultural heritage.

In the Asian region, where buildings predominantly constructed with wood are widely distributed, there is a pressing need for more economical and rapid long-term monitoring solutions. This context provides significant opportunities for the broader adoption of this model. Additionally, the characteristics of YOLO technology also endow it with the potential to be used to develop monitoring solutions for other types of architectural materials, such as stone, meeting the diverse conservation needs of various architectural heritages. This adaptability ensures that this technology can be tailored to the specific requirements of different types of building heritage, enhancing its applicability across a broad spectrum of conservation scenarios.

2. Wooden Structures in Fujian Tulou

2.1. Research Subject: Fujian Tulou

The Hakka people's migration patterns closely correlate with the geographical distribution of Tulou. Historically, the Hakka people have undergone a series of large-scale migrations, which have propagated the architectural form of the Tulou from Fujian to other regions [68]. Tulous are primarily found in the mountainous areas of Yongding, Nanjing, and Pinghe counties within Fujian Province [69]. These remote areas provided the Hakka people with relatively isolated and safe living conditions, which, in turn, promoted the development and perfection of Tulou architecture. As the Hakka people migrated outward, the architectural form of the Tulou also spread to other provinces. In Guangdong and Jiangxi, architectural forms similar to the Fujian Tulou emerged [70]. Although these Tulous are smaller in scale, they still retain the basic characteristics of the Fujian Tulou, such as an enclosed layout, rammed-earth walls, and wooden frames [3,71]. The appearance of

these Tulous reflects the Hakka people's inheritance and adaptation of traditional architectural forms during their migration. The discovery of some Tulou buildings in the Sichuan area is noteworthy. The "Huguang Fills Sichuan" policy enacted during the Qing Dynasty (A.D.1636–A.D.1912) may have contributed to this. During this period, the Sichuan authority resettled a large number of immigrants, including some Hakkas, from Hunan, Guangdong, and other places. These Hakka immigrants brought the architectural form of the Tulou to Sichuan, where they made appropriate modifications and adjustments according to the local geographical environment and building materials [72]. In addition to its domestic spread, the architectural form of the Tulou also spread to Southeast Asia with the Hakka people's overseas migration. In the late Ming and early Qing dynasties, many Hakkas began the migration activity of "going to Nanyang" to escape war and survival pressure [73]. They established new homes in Southeast Asia and brought the architectural concepts and techniques of the Tulou to the local area. Researchers have discovered Tulou-like buildings built by the Hakka people in countries like Malaysia and Indonesia (Figure 1).



Figure 1. Distribution of architectural works related to earthen houses (Tulou). (The small Chinese characters in the picture denote the country names originally on the base map).

The Hakka sought refuge in the mountains during periods of social unrest and conflict [74]. Tulous are known for their unique circular or square construction; these types of buildings aim to maximize defensive capabilities and promote public life [75]. The architectural structure of the Tulou mainly consists of a perimeter wall, housing units, a central courtyard, and defense facilities [76]. The perimeter wall is the main load-bearing structure of the Tulou. It is usually made of rammed earth or stone and has broad cantilevered eaves, tiled roofs, towering load-bearing rammed-earth walls, and a wooden frame structure [77]. The wall's thickness can reach 1.5–2 m. The perimeter wall's foundation penetrates deep into the ground and forms a tight connection with the base, creating a stable support system. The housing units are generally 3–5 stories high and are arranged in a circular or square pattern around the central courtyard, constructed from wooden frames and rammed-earth walls. Wooden beams and columns connect the housing units, forming

an overall circular or square structure [78]. The central courtyard is the Tulou's core space, as well as the main place for residents' activities and communication [79]. Typically, the courtyard is paved with stone slabs, and this courtyard also houses public facilities like wells and grain stores. The Tulou also has a variety of defense facilities, such as iron doors, gun holes, watchtowers, etc., to cope with foreign invasions and bandit harassment [80].

2.2. Analysis of the Wooden Structural Characteristics of Fujian Tulou

The use of wooden structures in the Tulou construction process reflects the wisdom of traditional Chinese architecture. During the construction of the earthen wall, builders would pre-embed wooden beams in the exterior wall, leaving space for doors and windows. This practice not only facilitates the subsequent installation of doors and windows but also creates an organic whole between the earthen wall and the wooden doors and windows. At the same time, window openings of different sizes create a natural and casual visual effect, reflecting the aesthetic pursuits of the builders. During the layered construction of the earthen walls, the builders would carve out recesses for the wooden floor beams at the top of each wall layer, and carpenters would erect large wooden columns, colloquially known as "shàngjià" ("上架"). The choice of time for column erection incorporates traditional customs of selecting auspicious dates, reflecting the builders' psychological desire for auspiciousness (Figure 2).



Figure 2. Aerial photos show cross and cantilever beams in Tulou ruins.

The method of supporting the floor reflects an organic combination of earth and wood structures: the outer side is supported on the load-bearing earthen wall, the inner side is supported on the column, beams and trusses are set in between, and the floor is laid on the rafters and reinforced with natural bamboo nails. At the same time, the end of the rafter supported on the earthen wall is slightly elevated to accommodate the height change after the earthen wall is air-dried and shrinks, showing the builders' profound understanding of material properties and their responsive measures. The roof of the Tulou adopts a column–beam wooden frame structure, similar to the practice in other traditional residences in China. The large quantities of wooden components used in mainstream Tulous also expose them to the risk of wood damage. The wooden structure necessitates long-term maintenance checks, timely observation, and fire protection for the wood materials. These structures have generated numerous safety hazards, and the overuse of the wooden structure's bearing capacity during influxes of large numbers of tourists [81] has also resulted in significant costs and safety issues for the protection and maintenance of Tulous.

In large-scale Tulous with a typical internal-corridor-containing round building structure, such as Huaiyuan Lou, the height of the building can reach up to four stories [82]. In addition to the outer earthen walls, the structures utilize wooden mortise-and-tenon frames or raised-beam hybrid frames [83]. The materials are mainly made from locally available wood, predominantly fir and pine. Tulou construction requires two kinds of timber: large timber and small timber. The main structure uses large timber, such as beams

and columns, typically made of fir. Large pine is essential for pilings in unstable foundations. For decorative purposes, such as embellishing partitions, floorboards, doors, and windows, small timber, such as imitation wood and boards, is typically used, along with fir. Pine, due to its durability, is ideal for stairs and floorboards. The large timber is prepared first, and then the small timber is prepared once the main body is complete. This type of wood has good texture, a uniform structure, and straight and long trunks, making it suitable for being processed into various planks or directly used as timber for beams and columns [84]. People have used surface carbonization through smoking to prevent dampness and enhance moisture resistance in well-preserved wood. Additionally, for some Tulous, caretakers have adopted the method of brushing tung oil on their wooden structures to reduce pest infestation and form a varnished surface to protect the wood itself.

In Tulou construction, wood is used for structure and space division. Once the earthen wall reaches a height of one story, the process of "qi zhu" (raising the column, "企柱") and beam erection commences. Wooden columns and beams ("kangshen" "扛身") are used to create a basic structure, while cantilever beams ("longgu" "龙骨") protrude inward from the column ends, supporting the bottom corridor while aligning with the top beam. To prevent sagging, the outer end of the cantilever beam is raised and placed in a groove within the wall. "Pengsheng" ("棚盛") runs parallel to the cantilever beam, with one end embedded in the wall and the other end resting on a beam slightly higher than the cantilever beam to maintain a level. From the second floor upwards, each floor has step columns (gallery columns, "廊柱") to support the cantilever beams. To maintain its level, carpenters raise the head of the cantilever beam. Carpenters must accurately measure the length and angle of the "kangshen" to ensure a precise connection when constructing a round Tulou. As the floors rise, the "jinzhu" (main pillar "金柱") moves inward for improved stability (Figure 3). In the "ziqiang" ("子墙") structure, no columns are erected; instead, cantilever beams are directly placed on the rammed wall, the "pengsheng" is placed inside the "ziqiang", and the gallery columns are positioned at the mouth of the cantilever beam's head. The floorboards are reserved for shrinkage, and the wedges gradually exit as the wall shrinks to keep the floorboards level. For a few Tulous, a "jinzhu" is erected in the "ziqiang", and brick columns are used instead of wooden columns on the ground floor in some Tolou [85].

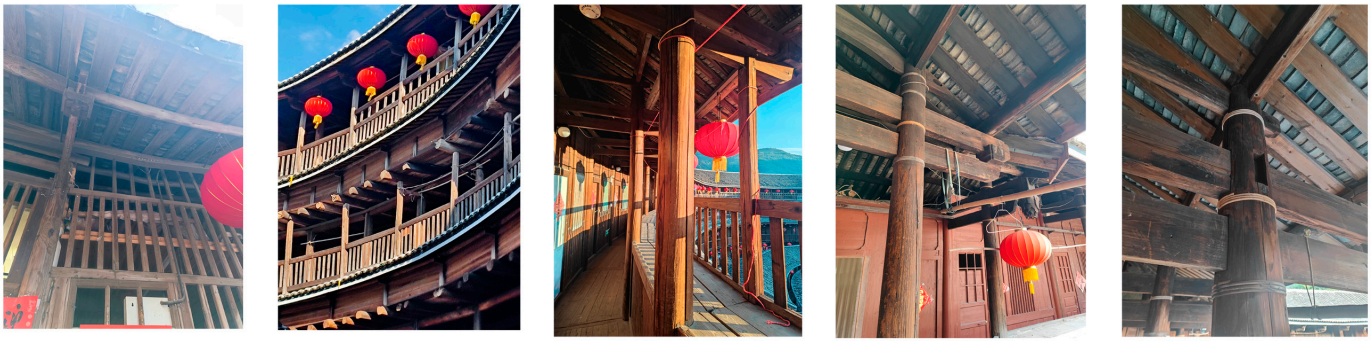


Figure 3. Tulou architecture showcases a variety of expressions of beams and columns.

In Tulous, wood not only serves a structural role but also helps in space division and floor connection. The roof's construction is an advanced stage in Tulou craftsmanship. The main beam is erected on top of the earthen wall, with the roof's slope and the length of the eaves precisely calculated. The roof uses "kang beams" ("扛梁"), requiring large timber, and a recessed design is made on the slope of the roof, referred to as "rangshui" ("让水"), giving the roof surface a curved appearance, enhancing its aesthetic appeal. The installation of "tongban" (bucket boards, "桶板") is an important step in roof construction. The arrangement of "tongban" is based on the length of the roof slope, which prevents the tiles from easily sliding off. The nailing method of "tongban" in square Tulous is relatively simple; however, in round Tulous, each piece of "tongban" requires precise measurement and

cutting considering its length and angle. Tulou roofs also have different constructions according to the shape of the corresponding building, such as hip or hanging-mountain roofs. The design of the sloping ridge and rain gutter needs to be both aesthetic and practical, ensuring the stability of the roof and its drainage function. After fixing the roof's "tongban" and tiles, the decoration of the wooden eave teeth enhances the beauty and stability of the eaves.

2.3. Analysis of Wood Damage Types and Contributing Factors in Fujian Tulou Structures

In Tulous, a traditional architectural form, wood is extensively used as the primary structural material. Despite centuries of natural erosion and human activity, these buildings often undergo renovation, even the well-preserved ones. Currently, Tulous face challenges like a decrease in residents, an increase in tourist traffic causing damage [86,87], and discontinuity in the transmission of repair techniques (Figure 4). The protection and restoration of the wooden structure are of imminent importance [88].

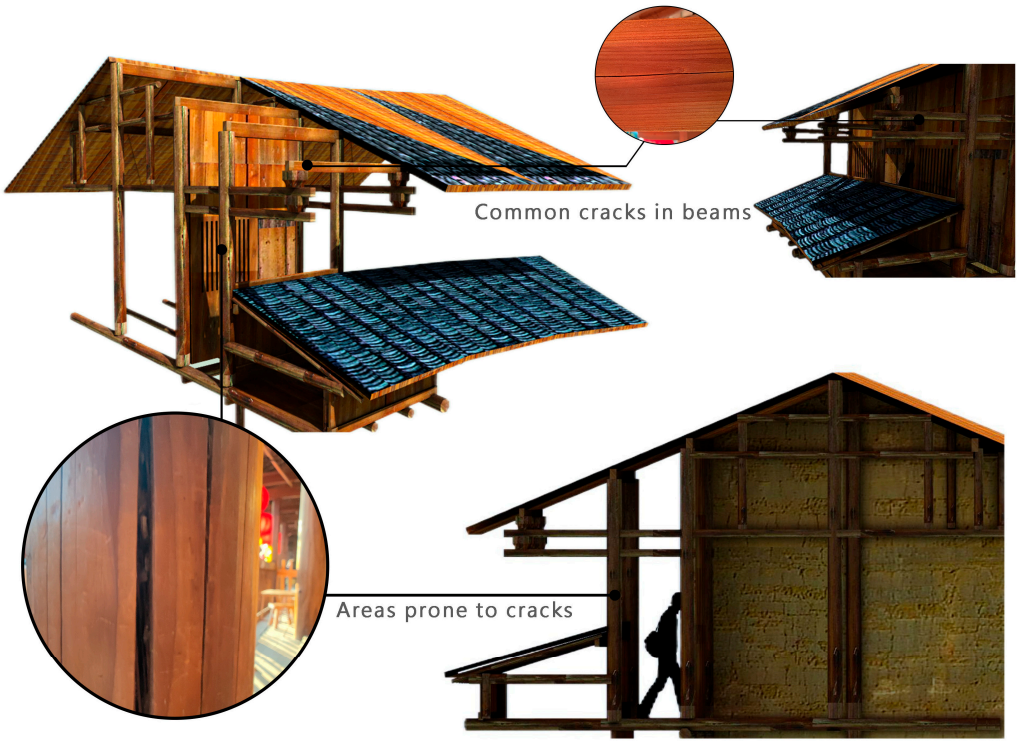


Figure 4. Common forms of damage in Tulous.

In traditional craftsmanship, wood is often treated via carbonization or painting to enhance its durability. However, influenced by commercialization and the decline of repair techniques, maintenance practices often lean towards cost-effectiveness, sacrificing the complexity of craftsmanship (Figure 5). A lack of adequate dehydration and pest-control treatment compromises the durability and structural integrity of wood, making it susceptible to damage. The climate in which Tulous are situated is characterized by high humidity and frequent rainfall, making the wood susceptible to moisture. Long-term sun exposure in the summer and high-humidity environments provide conditions for the growth of mold, leading to phenomena like cracking and holes in the wood, thereby shortening its lifespan. Current maintenance and repair techniques, to some extent, do not appropriately address the challenges posed by these natural and human factors. The loss of craftsmanship skills leads to the inability of repair projects to restore the original weather resistance and biological invasion resistance of the wood, further exacerbating the vulnerability of the wooden structure.



Figure 5. The site of rough wood craftsmanship processing.

We can summarize the three common types of damage by analyzing the causes: man-made stains, holes, and cracking (Figure 6).

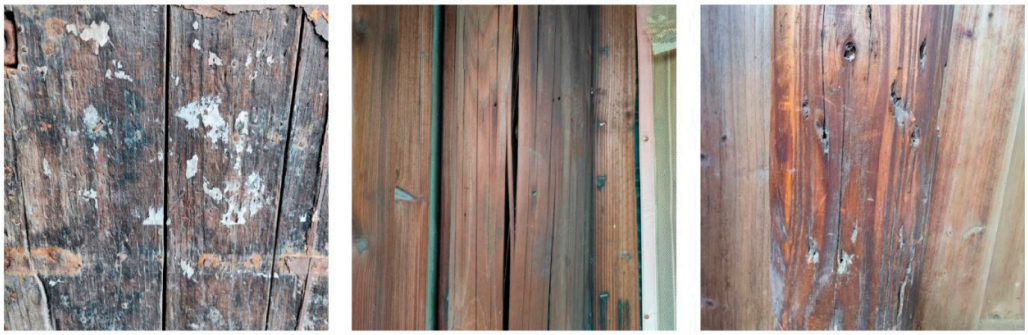


Figure 6. Types of damage to wooden materials collected on site.

(1) Stains. Stains on the surface of wood used in Tulous mainly stem from two causes: tourist activities and improper maintenance. Following the development of Tulou tourism, the resulting increase in the number of tourists has had a certain impact on the surface of the wood, leaving hard-to-remove stains. Also, food residue and beverage spills from tourists may leave stains on the wood surface. Apart from tourist factors, Tulous' daily maintenance and repair processes, if improperly conducted, can also lead to the formation of stains on the wood's surface. The use of inappropriate chemical cleaners, paints, or coatings may leave hard-to-remove stains or cause discoloration of the wooden surface. In the repair process, improper construction and the use of inappropriate materials affect the aesthetics of Tulou wood and may damage the wood over time (Figure 7).

(2) Holes. Over time, even treated wood's pest control and moisture resistance show a declining trend. Rainfall and temperature differences readily cause water seepage and condensation, which can lead to structural dampness. This dampness creates the ideal growth conditions for mold and insects, resulting in wood decay or hollowing, which, in turn, forms holes. Also, due to the increase in moisture, the toughness and solidity of the wood decrease, leading to an increase in holes created by building materials and reinforcements left during residents' and tourists' use and maintenance (Figure 7).

(3) Cracking. Improper use and lack of craftsmanship, long-term moisture caused by high humidity and frequent rainfall in the region where Tulous are situated, and brittleness caused by prolonged summer sun exposure can cause cracks in wood with proper moisture-proof performance and toughness under qualified craftsmanship and adequate

maintenance conditions. Mechanical damage due to increased tourist traffic, improper repair techniques, or the unreasonable use of tools on the wood itself accelerates its aging and breakage (Figure 7).

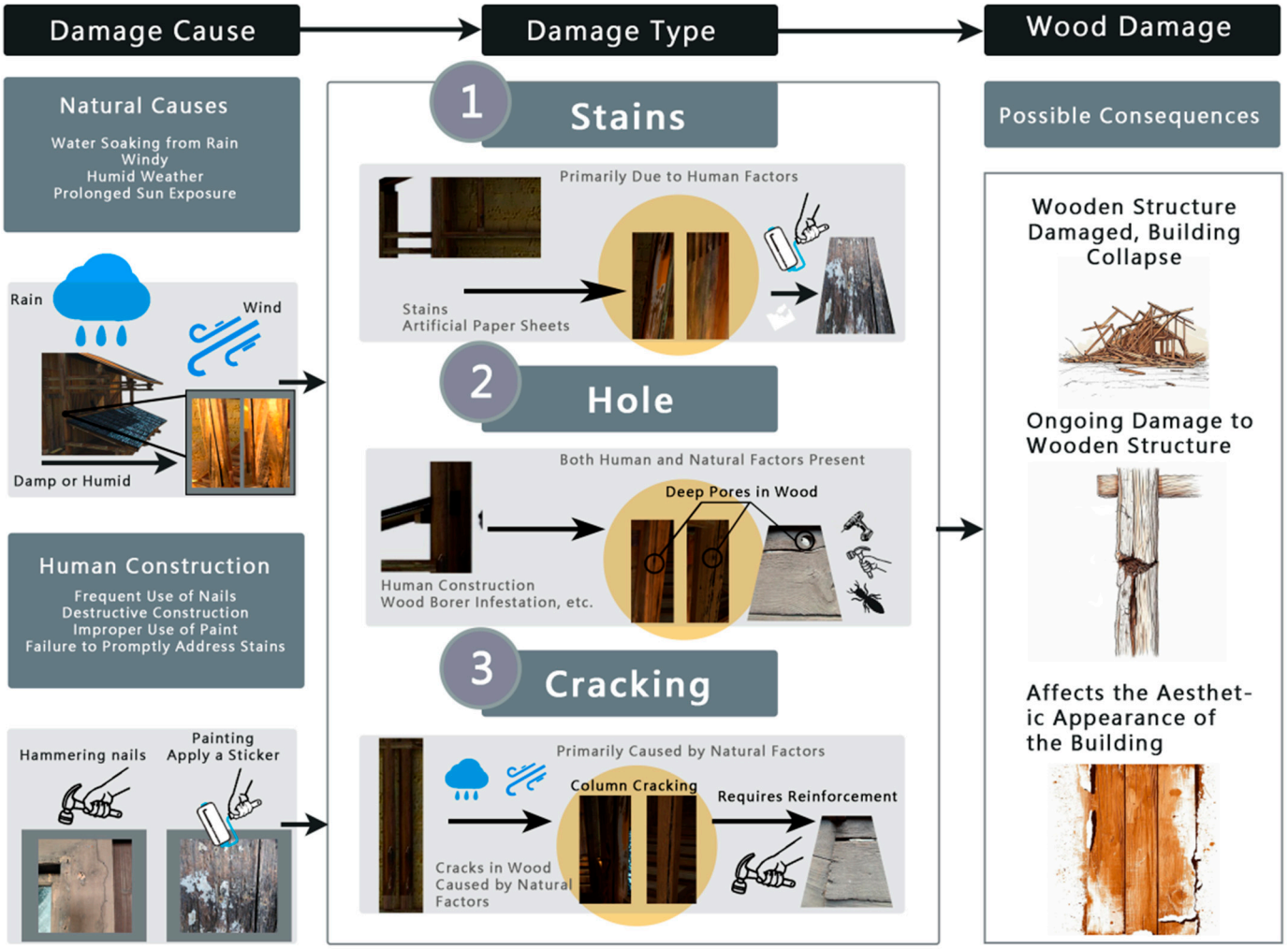


Figure 7. Wood damage types and contributing factors in Fujian Tulou structures.

3. Materials and Methods

3.1. Photographic Image Collection Source

We collected the samples for this study from the Gaobei and Nanxi Tulou clusters in southern Fujian (Figure 8), including ChaoYang Lou, ChengQi Lou, FuXing Lou, KuiJu Lou, QingCheng Lou, YanXiang Lou, and ZhenCheng Lou. In 2008, the UNESCO World Heritage List officially adopted these Fujian Tulous. Among them, ChengQi Lou and ZhenCheng Lou are some of the most representative and well-known Tulous in Fujian. We collected a total of 1723 image samples during the field investigation, and after screening for relevant features, we obtained 1506 valid samples. Among them, there were 45 image samples from ChaoYang Lou, 797 from ChengQi Lou, 21 from FuXing Lou, 288 from KuiJu Lou, 92 from QingCheng Lou, 186 from YanXiang Lou, and 77 from ZhenCheng Lou (please refer to Table A1 in Appendix B for the specific collection locations).



Figure 8. Researchers collecting images during fieldwork.

During the sample collection process, irregular wood grain and texture characteristics posed some challenges for the practical application of the samples. We utilized the DEMO method to refine and streamline the existing valid samples, taking into account the recognition rate of the original samples. Consequently, we optimized and adjusted the study to incorporate 300 valid samples for the image detection process. To ensure the representativeness of the sample set, we selected these 300 image samples from the previously mentioned 7 samples, based on their three specific damage features. The distribution of the samples was as follows: 9 from ChaoYang Lou, 80 from ChengQi Lou, 15 from Fuxing Lou, 71 from KuiJu Lou, 73 from QingCheng Lou, 42 from YanXiang Lou, and 10 from ZhenCheng Lou (please refer to Table A2 in Appendix B for the specific collection locations).

We selected locations within the buildings where damage features were clearly visible and lighting conditions were optimal, specifically with respect to the roof and corridors, as the sampling sites for the wooden components (Figure 9). We employed high-definition cameras to ensure the clarity of the samples. We manually categorized the damage features after collection, which led to the identification of three types of damage, as summarized above. We further refined the three types of damage, distinguished by different visual effects, after manual categorization. We then employed the YOLOv8 model to more accurately learn about and differentiate between these variations. This approach allows for precise damage detection and analysis in wooden structures in practical applications, providing technical support for historic building conservation and restoration.



Figure 9. Some of the collected images.

3.2. Research Process

In this investigation, we employed a rigorously tested and effective methodological approach, delving into the potential utility of computer vision technology in identifying damage to wooden components within Fujian Tulou, recognized as World Heritage Sites. The research encompassed the entire spectrum from data gathering to practical deploy-

ment. Within this framework, we developed and tested an automated method for detecting damage in wooden components using the YOLOv8 model. The objective of this study was to meld computer vision technology with the specific conservation needs of historical architecture, creating a model that can precisely and efficiently assess damage to wooden constructions. We designed this model to provide significant technological support for the long-term preservation of Fujian Tulou and other wood-based traditional buildings, with the goal of applying these advancements to real-world heritage preservation tasks. The research process is shown in Figure 10.

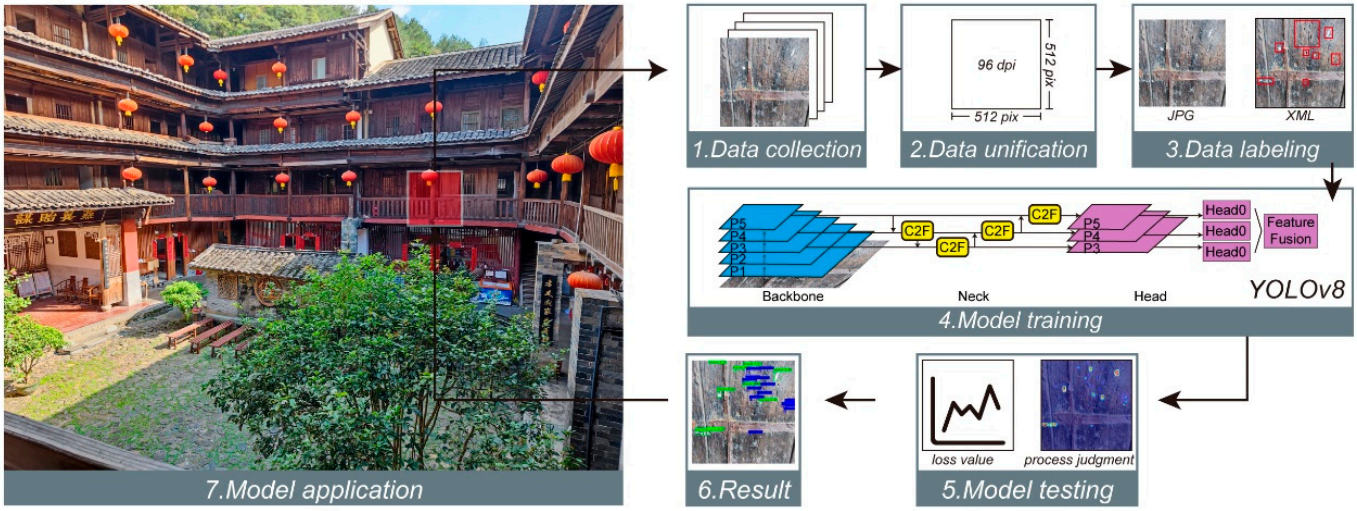


Figure 10. Research process.

(1) Material collection stage: The diversity of material sources and the representativeness of materials are critical for ensuring the model's stability and generalizability, which directly affect its practical utility. Therefore, during the material collection phase, we placed significant emphasis on acquiring a broad and representative range of images depicting damage to and human-induced stains on wooden structures. In typical Fujian Tulou locations, we collected 1506 high-definition images of various types of damage to wooden components. We collected images under different climatic conditions (sunny, overcast, and rainy) and varying lighting conditions (natural sunlight, shadows, and artificial light sources) to comprehensively cover variations observed in real-world scenarios. We aimed to capture the diverse manifestations of damage under various environmental conditions during a month-long data collection effort. Furthermore, to ensure that the collected data effectively reflected the characteristics of various types of damage, we focused not only on the diversity of damage types, such as water stains, surface scaling, color changes, and abnormal gaps, but also on the severity, size, shape, and color variations of the damage.

(2) Data-processing stage: We applied several image-preprocessing methods during the data preparation phase to improve image quality and standardize input data, thereby enhancing the model's training process. Then, we utilized techniques like histogram equalization to address issues related to varying lighting conditions and employed noise reduction methods to clarify features indicative of damage in the images. To facilitate processing by the YOLOv8 model, we standardized the images to a consistent size and resolution, resizing them all to 512 × 512 pixels. We adopted an overlapping tiling approach to ensure comprehensive representation of key features in the wooden structures within at least one image. This standardization helped stabilize the training process and lessen the computational load by eliminating discrepancies in image dimensions. Furthermore, we used techniques like rotating and flipping the images during the model training phase to increase the variety of the dataset, thereby improving the model's ability to generalize its robustness against varied inputs.

(3) Data annotation stage: The data annotation phase is a critical step in ensuring that the YOLOv8 model accurately learns features. Therefore, the research team committed to providing precise and uniform annotation data for each image. In this study, we used the professional image annotation tool, LabelImg, to accurately draw bounding boxes around various damage manifestations in the wooden structures depicted in the images. We assigned a unique identifier to each type of damage to establish a clear classification correspondence. Throughout the annotation process, this research emphasizes the consistency and accuracy of the annotations, with a dual-round review and adjustment procedure implemented to guarantee the precision of each bounding box and category label. Additionally, to enhance the model's generalization capacity, the annotation process includes damage at varying stages and degrees, ensuring the model can accurately predict damage of different severities.

(4) Model training stage: YOLOv8 was selected as the training model for the model training stage. The advantage of a machine-learning object detection model is that its unified network architecture can simultaneously locate and classify targets, enabling the model to achieve a high detection speed while maintaining high accuracy, making it ideally suited for application to soil construction damage in Fujian Tulou, as in our study. YOLOv8 is better than its predecessor in many ways, such as in terms of the model's structure, loss function, and training strategy. It was also given better feature pyramid networks (FPN) and path aggregation networks (PAN), which make it easier to combine features from different scales and find targets. We employed the following model-training strategy:

(a) Data sample division. To select 300 representative images, we further screened the 1506 sample images of damaged wood components collected in this study. Among them, there were 243 images in the training set, 27 images in the validation set, and 30 images in the test set, amounting to a total of 270 images used for the training and validation sets. We used the training set to train the model and continuously adjust its weight. We used the validation set to assess the model's performance and make necessary adjustments to the hyperparameters during the training process. After completing the training, we used the test set to finally evaluate the model's performance, ensuring it performed well on unseen data. We used the training and validation sets for cross-validation and other scenarios. When more training data were needed, the training and validation sets can also be used together for model training.

(b) Multi-stage model training. This model's training phase is divided into a frozen training phase and an unfrozen training phase. Only the last few layers of the network undergo training in the frozen training phase, while the parameters of the other layers remain unchanged, aiding in stabilizing the training process and preventing overfitting. During this phase, we used a larger batch size of 4, setting the initial learning rate to 0.01. The cosine decay (COS) strategy gradually reduces the learning rate as training progresses, bringing it down to a minimum of 0.0001. During the unfrozen training phase, we unlocked the parameters of all layers, enabling us to fully adjust the network to enhance model performance. To accommodate a more complex training process, this phase reduces the batch size to 2. We used a stochastic gradient descent (SGD) optimizer with a momentum of 0.937 to optimize the training process. Throughout the entire training process, we saved the model once per generation to allow us to resume training and conduct a subsequent analysis.

(5) Model testing stage: We conducted multiple tests on the model after training it to evaluate its performance and practical application effects. During the training process, the model generated a variety of loss values, including validation loss and training loss. These indicators help us to understand the performance of the model at different stages. To simplify the evaluation process, we used a simple mean average precision (mAP) calculation tool to track the mAP value during training so as to evaluate the detection accuracy of the model in real time during training. By analyzing the various loss values and mAP values during the training process, we selected the models with the best numerical performance

and the last generation of models for further testing, constituting a total of four models. We selected these models because they demonstrated high stability and accuracy during the training process, and we anticipated they would exhibit strong performance in actual tests.

During the model-testing phase, we used two key indicators, mAP (mean average precision) and lamr (log-average miss rate), to evaluate the numerical performance of the model. mAP is a performance indicator widely used in target detection tasks. It assesses a model's overall detection ability by calculating the average precision across all detection categories. A higher mAP value indicates that a model performs well in multiple detection categories. Lamr measures a model's detection performance by calculating the average error rate, particularly in high-false-positive cases. A lower lamr value indicates that a model has a lower error rate and higher reliability in the detection task. In addition to numerical tests, we also conducted actual image detection tasks to intuitively judge the effectiveness of the model. In this stage, the model detected wood structure damage (stains, cracks, and holes) in multiple test images. By observing the model's detection results, in this experiment, we were able to intuitively understand the model's performance in actual applications, including its recognition accuracy and detection speed for different types of damage. Finally, we selected the most suitable model for this study based on the models' performance regarding the above test indicators and their effectiveness in the detection tasks. We used this model for subsequent analysis and applications, such as the real-time monitoring of damage to wooden structures in Fujian Tulou and the formulation of maintenance recommendations.

(6) Results analysis stage: The objective of the model-testing phase was to conduct a thorough evaluation of its performance and pinpoint areas that may require further refinement. At the macro level, this analysis consolidates and examines the overall effectiveness of the model across the full test dataset, focusing particularly on its robustness and precision in the detection and localization of various types of damage. At the micro level, the focus is on how well the model can identify certain types of damage. This was determined by testing its performance in a variety of damage detection situations and examining things like detection rates, sensitivity, and any possible biases toward certain types of damage or conditions. By conducting a detailed examination of the model's operational behavior under and in various conditions and environments, respectively, this study highlights the challenges that the model might encounter in real-world applications and proposes strategies for its optimization. Additionally, this research explores instances of suboptimal model performance, investigating potential causes such as data imbalances and insufficient feature differentiation. These insights are crucial for guiding the further enhancement and fine-tuning of the model.

Through an in-depth investigation and exploration of these six steps, we developed a computer vision model capable of accurately identifying and locating damage within the wooden structures of Fujian Tulou (together constituting a World Heritage Site).

3.3. Model Settings and YOLOv8 Design

The network framework of the YOLOv8 model is shown in Figure 11. The architecture of this model consists of three main parts: the backbone network, the neck network, and the head network. These parts are responsible for feature extraction, feature fusion, and final detection prediction, respectively.

(1) The main function of the backbone network is to extract multi-scale feature maps from the input image. These feature maps provide a wealth of information for subsequent feature fusion and target detection. The specific steps are as follows: (a) An input image with dimensions of 512 × 512 × 3 is first processed by the backbone network. (b) A convolution operation in the P1 layer yields a feature map with dimensions of 256 × 256 × 80. This step extracts low-level features by reducing the resolution and increasing the number of channels. (c) In the P2 layer, a feature map with dimensions of 128 × 128 × 160 is obtained through further convolution and downsampling operations. This process extracts mid-level features. (d) In the P3 layer, a feature map with dimensions of 64 × 64 × 320 is

obtained through more convolution and downsampling operations. This process extracts deeper features and is suitable for small targets. (e) In the P4 layer, a feature map with dimensions of 32 × 32 × 640 is obtained through further downsampling operations. This process extracts higher-level features and is suitable for medium targets. (f) In the P5 layer, a 16 × 16 × 640 feature map is obtained through the final downsampling operation. This process extracts the highest level of features and is suitable for detecting large objects in an image. The model can capture information about objects of different scales using the above multi-scale feature maps, helping to improve detection accuracy.

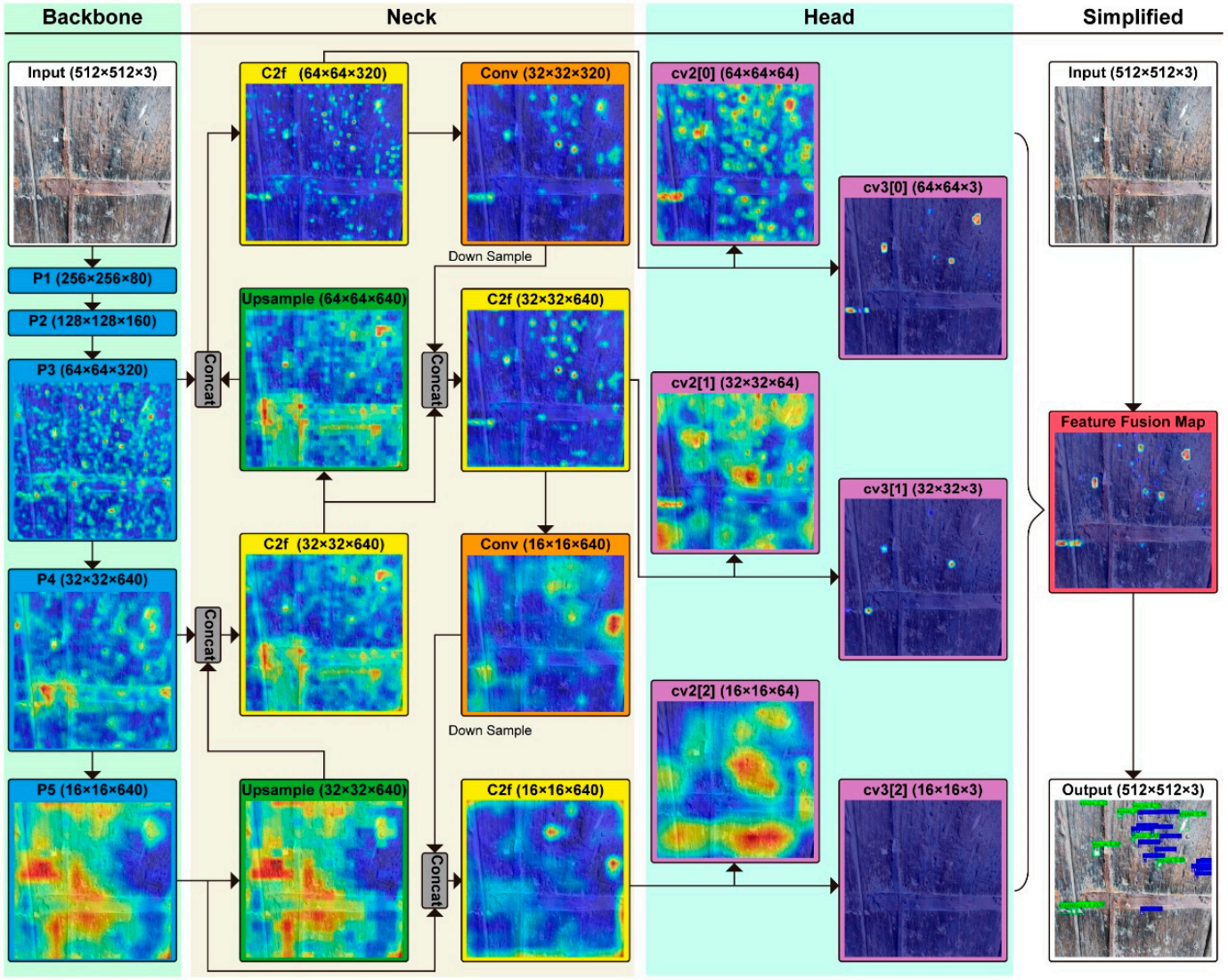


Figure 11. YOLOv8 model architecture.

(2) The neck network's primary function is to further process and fuse the multi-scale feature maps extracted by the backbone network. This step combines multi-scale features effectively using a feature pyramid network (FPN) and a path aggregation network (PAN), making it easier for the model to find objects of different sizes. The specific steps are as follows: (a) The P5 layer outputs a feature map, which undergoes a single upsampling procedure to yield a feature map with dimensions of 32 × 32 × 640. (b) We fuse the upsampled feature map with the P4 layer's feature map to create a feature map with dimensions of 32 × 32 pixels. This step combines high-level features with mid-level features to improve the robustness of detection. (c) After this fusion step, the feature map is upsampled again to obtain a feature map with dimensions of 64 × 64 × 640. (d) The upsampled feature map

is fused with the feature map of the P3 layer to obtain the final feature map with dimensions of 64 × 64 × 640. The neck network's multiple upsampling and feature fusion operations enhance the expression ability of multi-scale features and ensure effective object detection at different scales.

(3) The head network is mainly responsible for processing the fused feature map to generate the final detection result. The specific steps are as follows: (a) A convolution operation processes the 64 × 64 × 640 feature map, resulting in a 32 × 32 × 320 feature map. (b) Downsampling is performed again to obtain a feature map with dimensions of 16 × 16 × 640, and high-level features are gradually extracted through multiple convolution operations. (c) After further processing, the generated feature map yields a detection result with dimensions of 512 × 512 × 3. Through this design, the head network can make predictions at multiple scales, ensuring the effective detection of objects of different sizes.

Overall, the YOLOv8 model used in this study does a good job of turning the input image into target detection results using the backbone network, the neck network, and the head network. The backbone network can extract information from multi-scale features to ensure comprehensive coverage of targets of different sizes. The neck network can perform multiple feature fusion and upsampling operations, enhancing the model's multi-scale feature expression ability. The head network can make predictions at multiple scales, ensuring the accurate detection of targets of varying sizes. The design of these architectures is suitable for detecting damage (stains, cracks, and holes) to wooden structures in Fujian Tulou. Through multi-scale feature extraction and fusion strategies, the model can efficiently and accurately identify damage of different types and scales, providing strong technical support for the protection and restoration of Fujian Tulou.

4. Results: Automatic Recognition Result Analysis

4.1. Model Experiment Process

In this study, we conducted four experiments in order to evaluate the performance of the YOLOv8 model in the damage detection of wooden structures in Fujian Tulou, with different settings applied and adjustments made for each experiment.

4.1.1. Test Results of Experiment 1: Preliminary Testing

We used an initial baseline setting in the first experiment. We set the model parameters and dataset-partitioning values to the default settings to allow them to serve as comparison benchmarks for subsequent experiments. The primary goal for this experiment was to build a basic model and determine the direction of improvement through preliminary result analysis. A look at important numbers like average precision, log-average miss rate, F1, recall, and precision (as shown in Figure 12) shows how we were able to determine how well the model could find different kinds of damage.

(a) For the hole label, the average precision is 0.51, which means that among all hole detections, the correct detection rate is more than half, but it is not high. At the same time, the log-average miss rate is 0.73, indicating that the model has a serious problem of missing holes when detecting holes. The F1 value is 0.25, which is a low comprehensive indicator, indicating that the model performs poorly in balancing precision and recall rate. Specifically, the recall rate is 0.59, indicating that the model can detect about 59% of the holes, but its precision is only 0.14, which means that in regard to the number of holes detected, only 14% are real holes, indicating a high false-positive rate.

(b) For the stain label, the average precision is 0.39, slightly lower than that of the hole label, but the log-average miss rate is higher, amounting to 0.85, which means that there is a greater risk of missed detection in stain detection. Nevertheless, the F1 value is 0.4, which is relatively high, reflecting a satisfactory balance between precision and recall. The specific recall rate is as high as 0.85, showing that the model can identify most stains, but its precision is 0.3, which still indicates a certain degree of false detection.

(c) For the cracking label, the average precision is the lowest, amounting to only 0.31, indicating that the model has poor accuracy in crack detection. The log-average miss rate is the highest, at 0.88, indicating that the model misses the most when detecting cracks. The F1 value is 0.25, similar to that of the hole label, indicating that the model also performs poorly in this regard. The recall rate is 0.52, meaning that the model can only identify about half of the cracks, while the precision is 0.16, showing a high false-positive rate for crack detection.

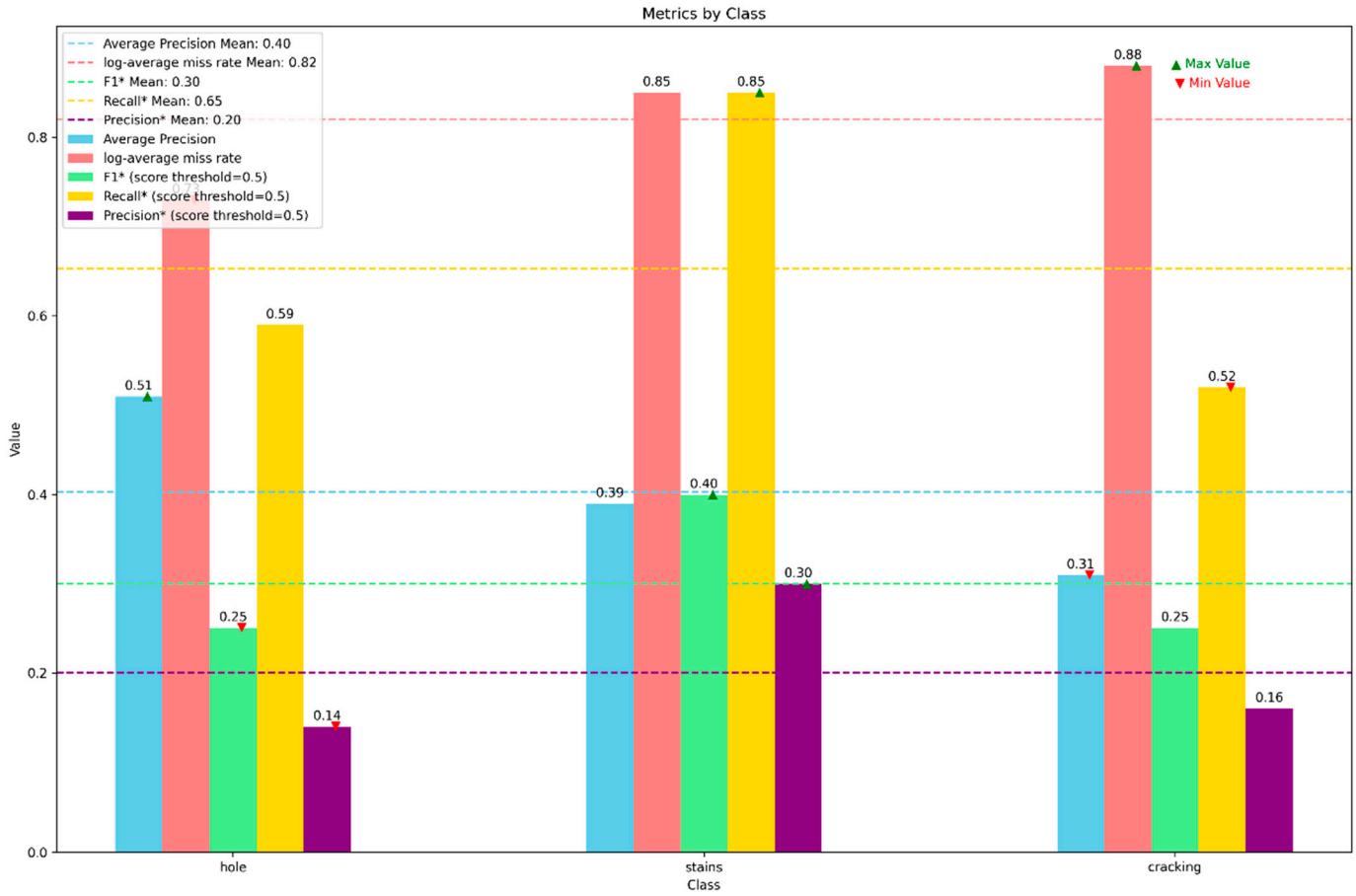


Figure 12. The results show the various indicators of the model in Experiment 1 (asterisk * in the figure indicates the median). In the figure, F1* indicates score threshold = 0.5; Recall* indicates score threshold = 0.5; Precision* indicates score threshold = 0.5.

In the confusion matrix diagram (Figure 13), we can see the classification performance of the model for different labels (holes, stains, and cracks) and backgrounds. The confusion matrix is presented in a normalized form, and the value in each cell represents the matching ratio between the predicted category and the actual category, thus revealing the accuracy and false-alarm rate of the model for different labels.

(a) Models with cracked labels performed poorly. Specifically, when the actual label is "cracked", the model only correctly predicts cracking 16% of the time, misclassifies it as background 39% of the time, and misclassifies it as a hole 10% of the time. This shows that the model has a lot of difficulty distinguishing splits from background, and its false-alarm rate is high, which is a significant bottleneck in the current model's performance. Furthermore, the 10% false-positive rate demonstrates that the model has some difficulty distinguishing cracks and holes.

(b) The model with stained labels performed slightly better. When the actual label is a stain, the model has a 58% probability of correctly predicting it as such, reflecting the model's relatively high accuracy in detecting stains. However, the probability that the

model misjudges stains as background is as high as 50%, showing that the model still has significant confusion problems when dealing with the distinction between stains and background. This high false-positive rate not only affects the model's reliability in stain detection but also shows that the background has a greater impact on the model's overall performance.

(c) The model performance for hole labels is moderate. Specifically, when the actual label is a hole, the model correctly predicts it as such 38% of the time but misclassifies it as background 52% of the time and misclassifies it as a crack 11% of the time. These high false-alarm and low correct prediction rates indicate that the model needs further optimization for hole detection. In particular, the 52% false-alarm rate shows that the model has great difficulty distinguishing holes from the background.

(d) The background false-alarm rate is generally high, significantly affecting the model's overall performance. Cracked, stained, or hole labels significantly increased the background false-alarm rate. For example, when the actual label is cracked, 84% of the samples are misclassified as background; when the actual label is stained, 42% of the samples are misclassified as background; and when the actual label has holes, 52% of the samples are misclassified as background. These data show that background categories more greatly interfere with model classification and are one of the key factors affecting model accuracy and reliability.

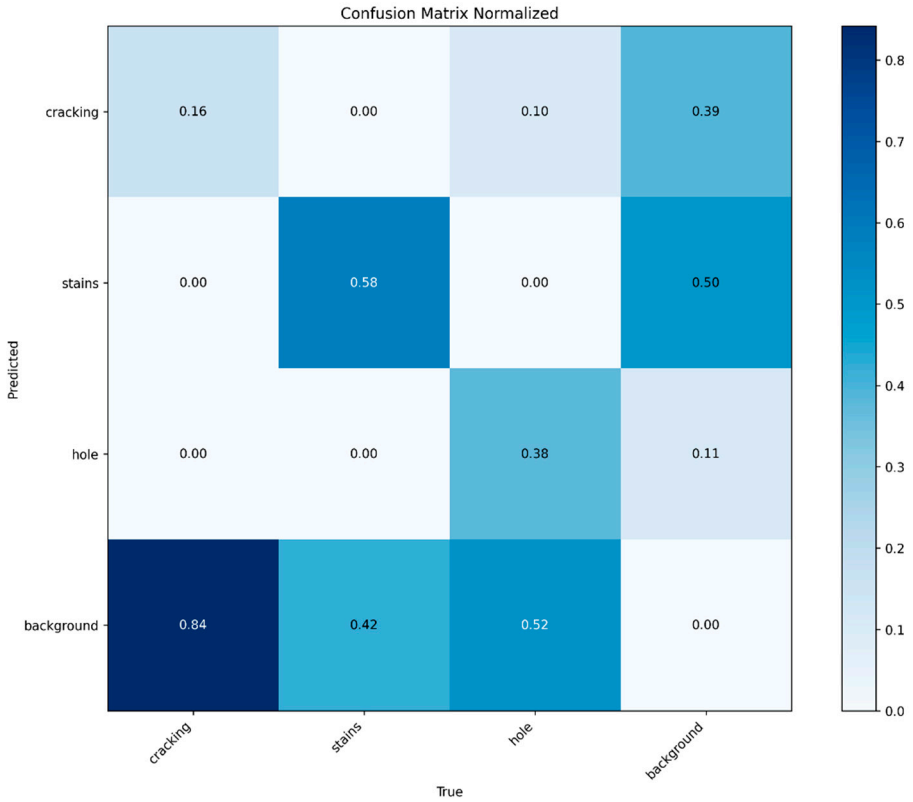


Figure 13. The confusion matrix pertains to the models in Experiment 1.

Therefore, the first experiment reveals the difference in model performance between different label classifications. Although the model has high accuracy when detecting stains, it performs poorly when detecting cracks and holes. In particular, the high false-alarm rate in the background significantly affects the overall performance of the model.

4.1.2. Test Results of Experiment 2: Removal of Complex Background and Optimization

In the second experiment, we noticed that some samples' backgrounds were too complex, interfering with the model's feature extraction process and leading to poor detection results. To solve this problem, we deleted samples with exceedingly complex back-

grounds and finally screened out 300 samples (originally 1506) to ensure that the remaining samples had clearer foreground targets. After this adjustment, the model's accuracy improved significantly (as shown in Figure 14), and this study confirmed the importance of dataset quality for model performance. The following is a detailed analysis of the experimental results:

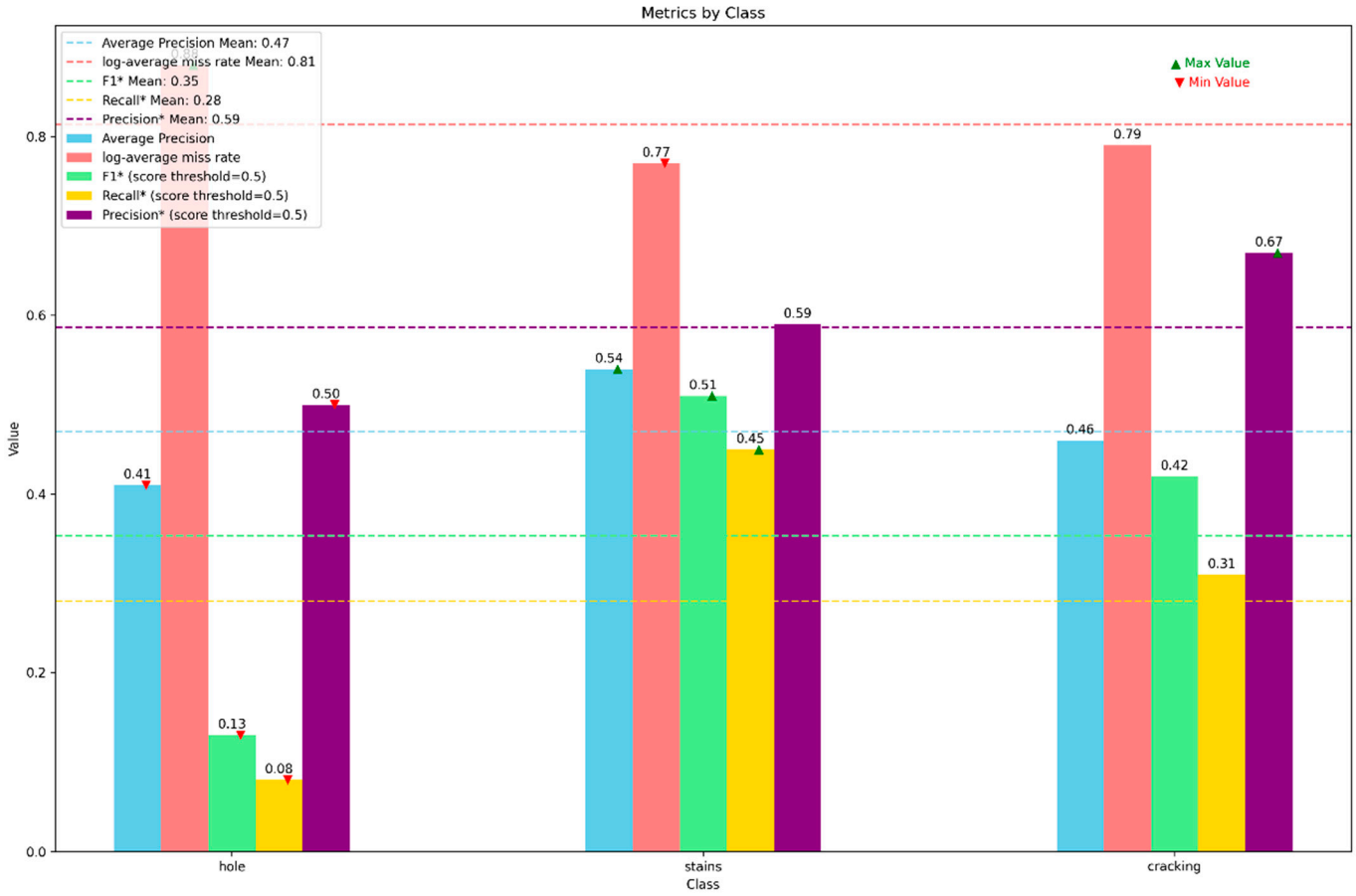


Figure 14. The results show the various indicators of the model in Experiment 2 (asterisk * in the figure indicates the median). In the figure, F1* indicates score threshold = 0.5; Recall* indicates score threshold = 0.5; Precision* indicates score threshold = 0.5.

(a) In this experiment, the average accuracy for the hole label dropped to 0.41, which was lower than the first experiment's value of 0.51, indicating that the model's accuracy in detecting holes decreased. The log-average miss rate increased to 0.88, showing that the missed detection problem is more serious. The F1 value is 0.13, which is significantly lower than the previous value of 0.25, meaning that the overall performance deteriorated. The recall is only 0.08, which is significantly lower than the value of 0.59 in the first experiment, indicating that the model misses most holes. The precision increased significantly, reaching 0.5, indicating that the model is highly accurate in detecting most holes, but some labels' accuracy is low, resulting in a decrease in average accuracy. This result demonstrates that there is a certain degree of error in the model's hole label samples.

(b) The average accuracy for the stain label significantly improved to 0.54 from 0.39 in the first experiment, demonstrating the enhanced accuracy of the model in stain detection. The log-average miss rate significantly improved to 0.77 from the previous value, 0.85. The F1 value improved to 0.51 from the previous value of 0.4, demonstrating a better balance between precision and recall. The recall rate dropped slightly to

0.45, but the precision increased to 0.59, indicating that the model's accuracy in stain detection increased and its false detection rate decreased.

(c) For the cracking label, the average accuracy increased to 0.46, a significant improvement compared to the value of 0.31 shown in the first experiment, showing that the model's accuracy in crack detection improved. The log-average missed detection rate dropped to 0.79, a decrease from the previous value of 0.88. The model's overall performance greatly improved, as evidenced by the F1 value increasing to 0.42, significantly higher than the value of 0.25 in the first experiment. Although the recall rate decreased to 0.31 from the previous value of 0.52, the precision significantly improved, reaching 0.67, demonstrating a significant improvement in the model's crack detection accuracy and a reduction in the false detection rate.

In terms of the confusion matrix plot (Figure 15), for the cracking label, the correct classification rate is 0.41, which means that the model can correctly detect 41% of the cracked samples. The false detection rate is high, especially for the background label; it can be seen that 0.26 cracked samples were mistaken for background, showing that the model has significant difficulty in distinguishing cracks from the background.

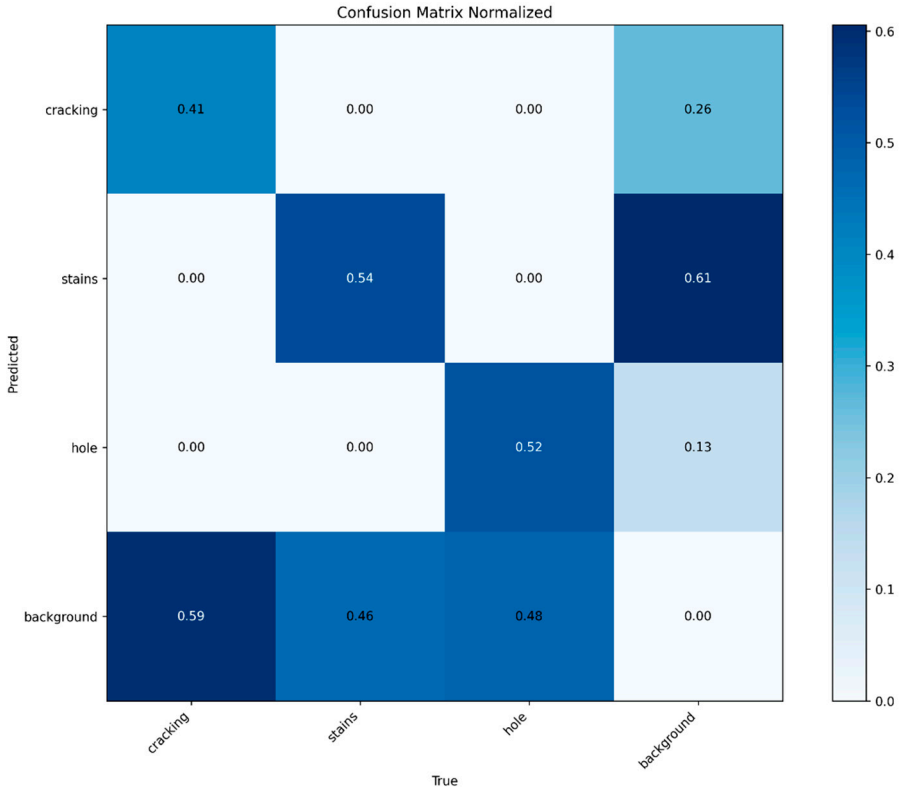


Figure 15. The confusion matrix pertaining to the Experiment 2 model.

(a) For the stains label, the correct classification rate is 0.54, and the model can correctly identify 54% of the stain samples. However, the model misclassifies 0.61 of the stain samples as background, suggesting significant room for improvement in stain detection, particularly in regard to distinguishing stains from the background.

(b) For the hole label, the correct classification rate is 0.52, indicating that the model performs relatively well in detecting holes. However, some holes were still misclassified as background (0.13), showing that the model occasionally misclassifies holes as background when detecting holes, but the false-positive rate is low.

(c) The false-positive rate of the background is still high: 59% of background images were classified as cracks, 46% were classified as stains, and 48% were classified as holes.

This shows that the model has obvious deficiencies in distinguishing the background from other labels and needs further optimization.

Taken together, the results of the second experiment show significant improvements in stain and crack detection, especially in terms of accuracy. Although the recall rate declined, the overall performance is more balanced. The performance in hole detection regressed, especially with respect to the significant drop in the recall rate, indicating that the model's missed detection problem for this task is more serious and that further optimization is needed to improve its overall detection ability.

4.1.3. Test Results of Experiment 3: The Optimization of Label Quality Again

In the third experiment, we ameliorated the label quality problems found in the first two experiments. As a result of the phenomenon of missing or mislabeling a small number of labels during the data-labeling process, the features learned by the model were inaccurate, affecting the overall detection performance. To solve this problem, we strictly reviewed and relabeled all the samples to ensure the labels' accuracy. The following is a detailed analysis of the experimental results (Figure 16).

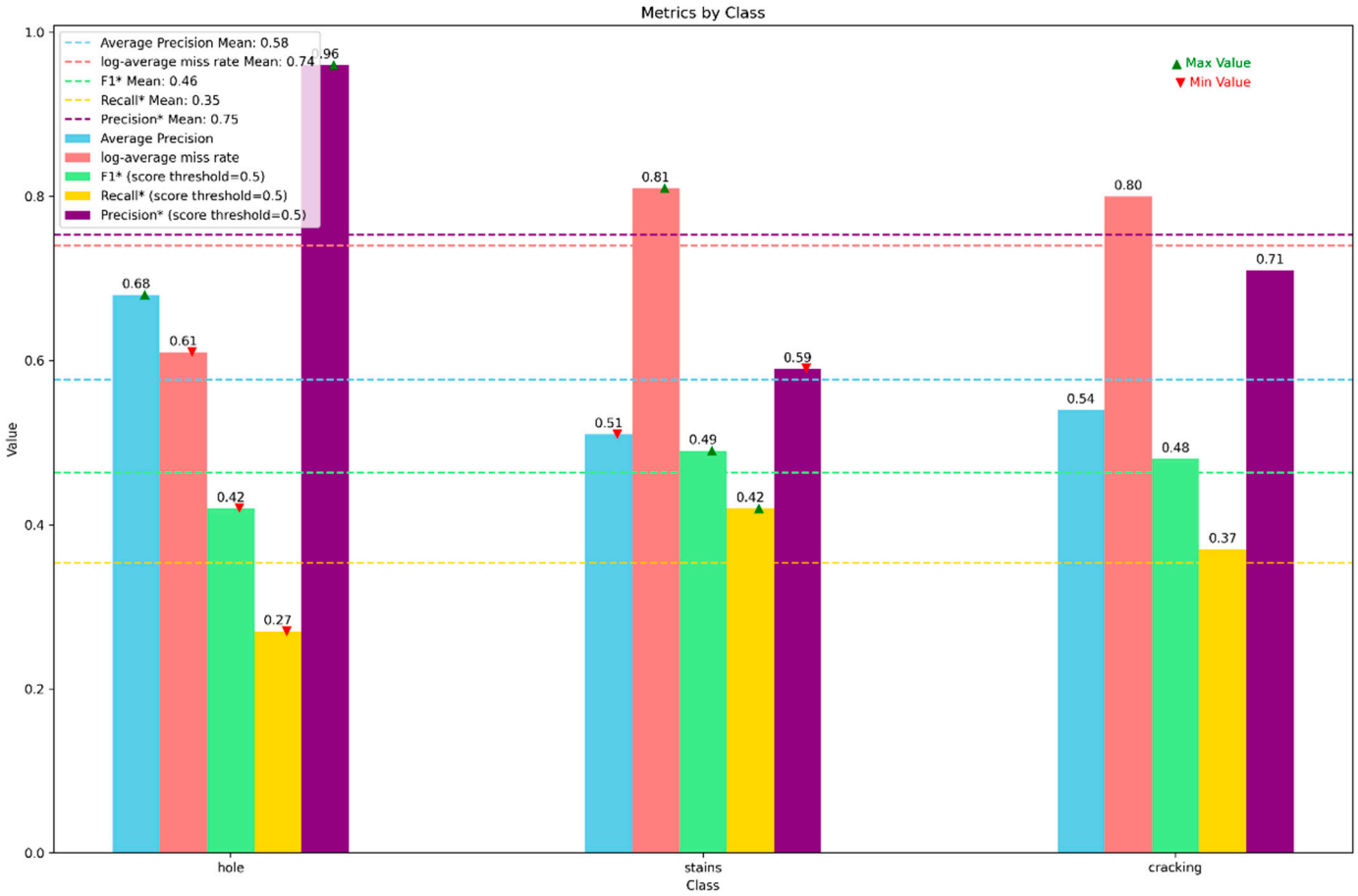


Figure 16. The results show the various indicators of the model in Experiment 3 (asterisk * in the figure indicates the median). In the figure, F1* indicates score threshold = 0.5; Recall* indicates score threshold = 0.5; Precision* indicates score threshold = 0.5.

(a) The average accuracy for the hole label significantly improved from 0.41 in the second experiment to 0.68, demonstrating a significant improvement in the model's hole detection accuracy. The log-average miss rate significantly improved from the previous value of 0.88 to 0.61, signifying a reduction in missed detections. The F1 value increased to 0.42, showing an improvement in overall performance. The recall rate was 0.27. Although this recall is low, it is a significant improvement from the previous

value of 0.08, demonstrating the model's ability to identify more holes. The model's precision significantly increased to 0.96, demonstrating its exceptional accuracy in hole detection and a significant reduction in the false detection rate.

(b) For the stains label, the average precision is 0.51, which is slightly lower than the value of 0.54 in the second experiment but still higher than the value of 0.39 in the first experiment, indicating that the model has good stability in stain detection. The log-average miss rate was 0.81, which is slightly higher than the previous value of 0.77, indicating that the problem of missed detection still exists. The F1 value was 0.49, which is slightly lower than the previous value of 0.51, but it is still at a high level overall. The recall was 0.42, slightly lower than the previous value of 0.45, but the precision remained at 0.59, indicating that the accuracy of the model in stain detection is still high.

(c) The average accuracy for the cracking label improved to 0.54 from 0.46 in the second experiment, suggesting an enhancement in the model's crack detection accuracy. The log-average miss rate was 0.80, slightly lower than the previous value of 0.79, showing an improvement in missed detections. The F1 value increased to 0.48, a significant improvement from the previous value of 0.42, showing an improvement in overall performance. The model's recall improved to 0.37 from the previous value of 0.31, demonstrating its ability to identify more cracking situations. The precision is 0.71, which is significantly higher than the previous value, 0.67, indicating that the accuracy of the model in crack detection has been further improved.

Figure 17 shows the confusion matrix plot. For the cracking label, the correct classification rate is 0.67. This means that the model can correctly identify 67% of the cracked samples, constituting a significant improvement. However, the model still misclassifies 0.30 of the cracked samples as background, indicating difficulty in distinguishing cracks from the background in some cases. Although these results show improvement, the model still needs further optimization.

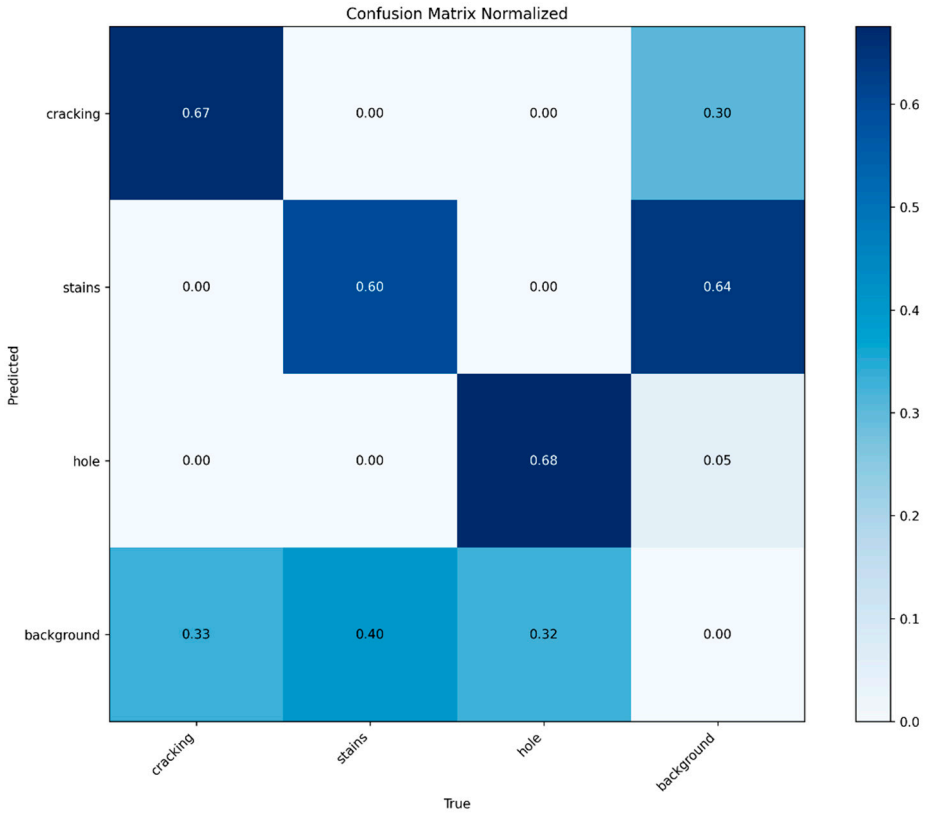


Figure 17. The confusion matrix pertaining to the Experiment 3 model.

(a) For the stains label, the correct classification rate is 0.60, i.e., the model can correctly identify 60% of the stain samples, which is a relatively stable performance. However, the model misclassifies 0.64 of the stains samples as background, suggesting there is significant room for improvement in distinguishing stains from the background. This false-positive rate is relatively high, affecting the overall detection performance of the model.
(b) The model successfully identified most hole samples, with a correct classification rate of 0.68 for the hole label. The model effectively distinguishes holes from the background, as evidenced by the low false-positive rate and the fact that it misclassified only 0.05 of the hole samples as background.
(c) The background false positives showed a slight decrease compared to the previous two experiments: 33% were classified as cracks, 40% were classified as stains, and 32% were classified as holes. These results show that the model still has significant shortcomings in distinguishing the background from other labels.

In conclusion, the third experiment's results demonstrate a significant improvement in the model's ability to detect holes and cracks, particularly in terms of accuracy, but there is still room for improvement in the recall rate. The stain detection performance remained stable, with little variation in precision and recall. Overall, the model's performance on each label improved to varying degrees, but it still requires further optimization in terms of its missed detection rate and recall rate to enhance its comprehensiveness and reliability of detection.

4.1.4. Test Results of Experiment 4: Hyperparameter Tuning

In the fourth experiment, the hyperparameters were adjusted based on the previous experiments' results. We primarily adjusted the key parameters, including the learning rate, momentum, and batch size. Specifically, we changed the initial learning rate from 0.01 to 0.001, the minimum learning rate from 0.0001 to 0.00001, the momentum coefficient from 0.937 to 0.9, the batch size from 4 to 8 in the frozen training phase, and the batch size from 2 to 4 in the unfrozen training phase to enhance the model's convergence speed and detection accuracy. The following is a detailed analysis of the experimental results (Figure 18).

(a) For hole labels, the average accuracy is 0.66, which is slightly lower than the value of 0.68 in the third experiment but still a high level. The log-average miss rate is 0.63, slightly higher than the previous value, 0.61, indicating that the number of missed detections increased slightly. The F1 value increased to 0.55, showing a significant improvement in overall performance. The recall rate is 0.42, an improvement from 0.27 in the third experiment, indicating that the model can identify more holes. The precision is 0.82, which is slightly lower than the previous value, 0.96, but still high, indicating that the model has high accuracy when detecting holes.
(b) For the stains label, the average precision is 0.56, which is an improvement from 0.51 in the third experiment, indicating that the model has improved its accuracy in stain detection. The log-average miss rate is 0.78, which is a decrease from the previous value, 0.81, indicating that the problem of missed detection was ameliorated. The F1 value increased to 0.53, which is an improvement from the previous value, 0.49, showing a better balance between precision and recall. The model's accuracy in stain detection further improved to 0.44, slightly higher than the previous value, 0.42, and its precision increased to 0.67, indicating a reduction in the false-positive rate.
(c) For the cracking label, the average precision is 0.50, which is lower than the value of 0.54 obtained in the third experiment, indicating that the model's accuracy in crack detection decreased. The log-average miss rate is 0.85, slightly higher than the previous value, 0.80, indicating that the number of missed detections increased. The F1 value is 0.48, the same as before, showing that the overall performance of the model in crack detection remains stable. Recall is 0.37, the same as before, and precision is

0.71, which is also unchanged, showing that the model's accuracy in crack detection remains stable.

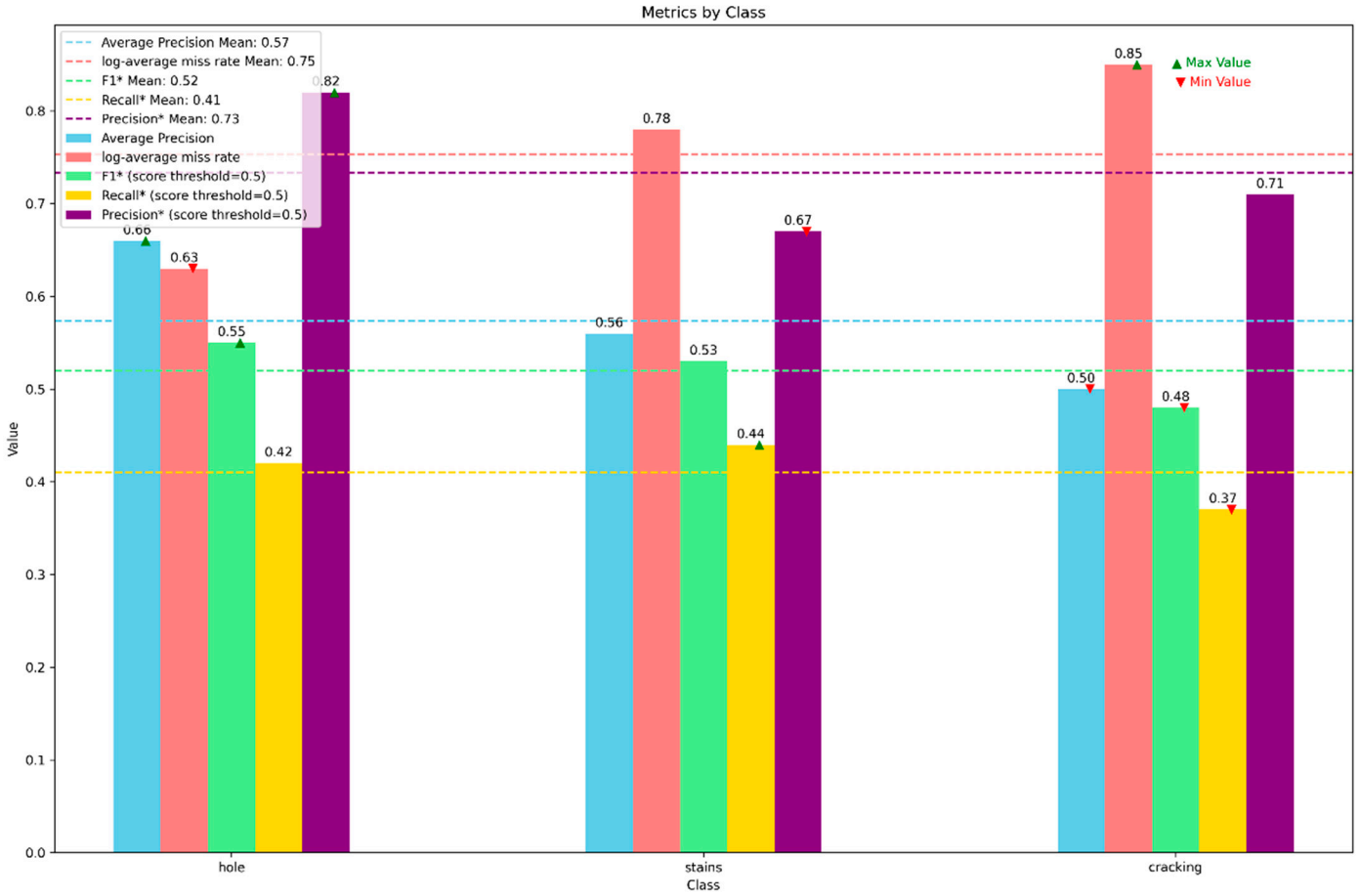


Figure 18. The results show the various indicators of the model in Experiment 4 (asterisk * in the figure indicates the median).In the figure, F1* indicates score threshold = 0.5; Recall* indicates score threshold = 0.5; Precision* indicates score threshold = 0.5.

The confusion matrix diagram (Figure 19) shows that the correct classification rate for the cracking label is 0.57, which means that the model can correctly identify 57% of the crack samples. This is less than the value in the previous experiment. The model misclassified 35% of the cracking samples as background, showing that the model still has some difficulty distinguishing between cracks and background and needs further optimization.

(a) For the stains label, the correct classification rate is 0.54, which is the same as that observed in the previous experiment, showing the stability of the model in stain detection. However, the model still requires improvement in terms of distinguishing stains from background, as it misclassified 45% of the stain samples as background. Although the correct identification rate for stain detection is high, the background false detection rate affects the model's overall performance.

(b) For the hole label, the correct classification rate is 0.62, showing that the model performs well in hole detection. Although it is a low value, the model misclassifies 20% of the hole samples as background, indicating a certain error in distinguishing holes from the background.

(c) Compared to the previous experiment, the background false-alarm rate showed a slight increase: 43%, 46%, and 38% of the samples were classified as cracks, stains, and holes, respectively. This demonstrates the clear deficiencies in the model's ability to distinguish background from other labels, highlighting the need for further optimization in this area.

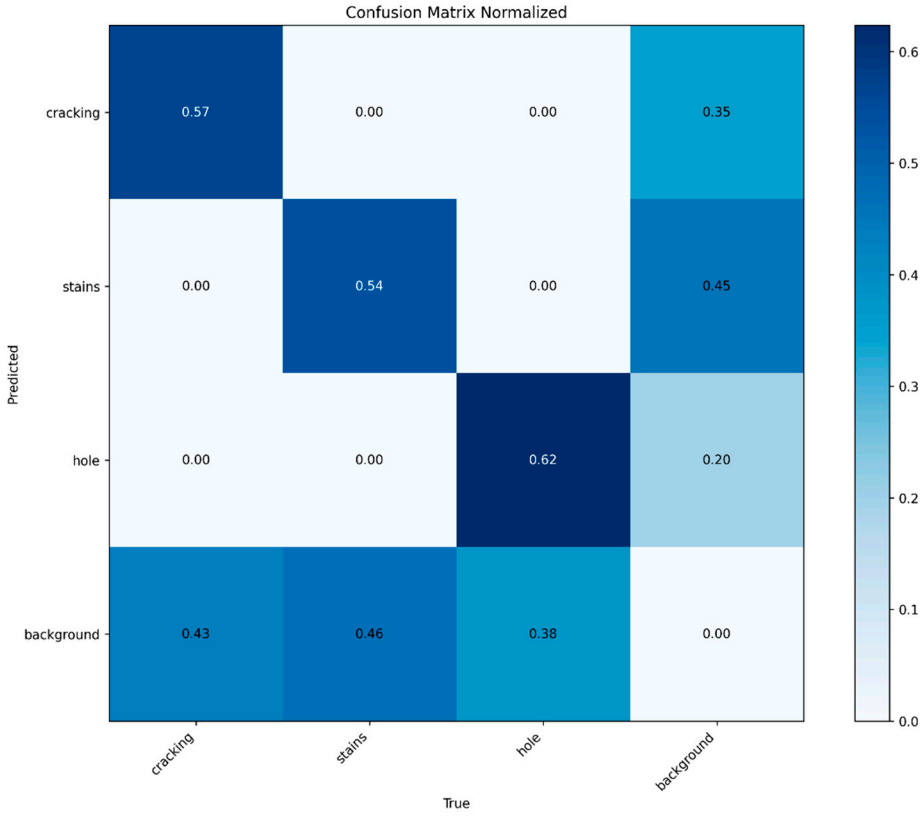


Figure 19. The confusion matrix pertaining to the Experiment 4 model.

In summary, the results of the fourth experiment show that the model's hole and stain detection rates are similar to those obtained in the previous experiment, with slight improvements in recall and precision. Although its accuracy in crack detection decreased, its overall performance remained stable. While the model's performance on each label improved to varying degrees, we still need to further optimize the missed detection rate and recall rate to enhance the comprehensiveness and reliability of detection.

The model's performance on each label underwent distinct changes and improvements, as evidenced by the four experiments mentioned. Considering multiple indicators such as average accuracy, F1 value, recall rate, and precision, the model performance values of the third and fourth experiments are relatively close. However, the fourth experiment's F1 value and recall rate are higher overall, and the model has the best overall performance in stain and hole detection. Considering the comprehensive performance exhibited for each label, especially the balance between F1 value and recall rate, it can be said that the model from the fourth experiment performs best in detecting holes and stains while remaining stable in crack detection. As a result, in the following study, we will use the fourth experiment's model as the basis for further research and analysis.

4.2. Loss during Training

In this experiment, we significantly improved the model's overall performance in terms of various types of damage detection through strict label quality control and relabeling. The overall mAP reached 57%, performing relatively well in various types of damage detection (Figure 18). Therefore, we selected this model as the basis for further research. Figure 20 depicts the YOLOv8 model's changes in training loss, validation loss, and mAP during training.

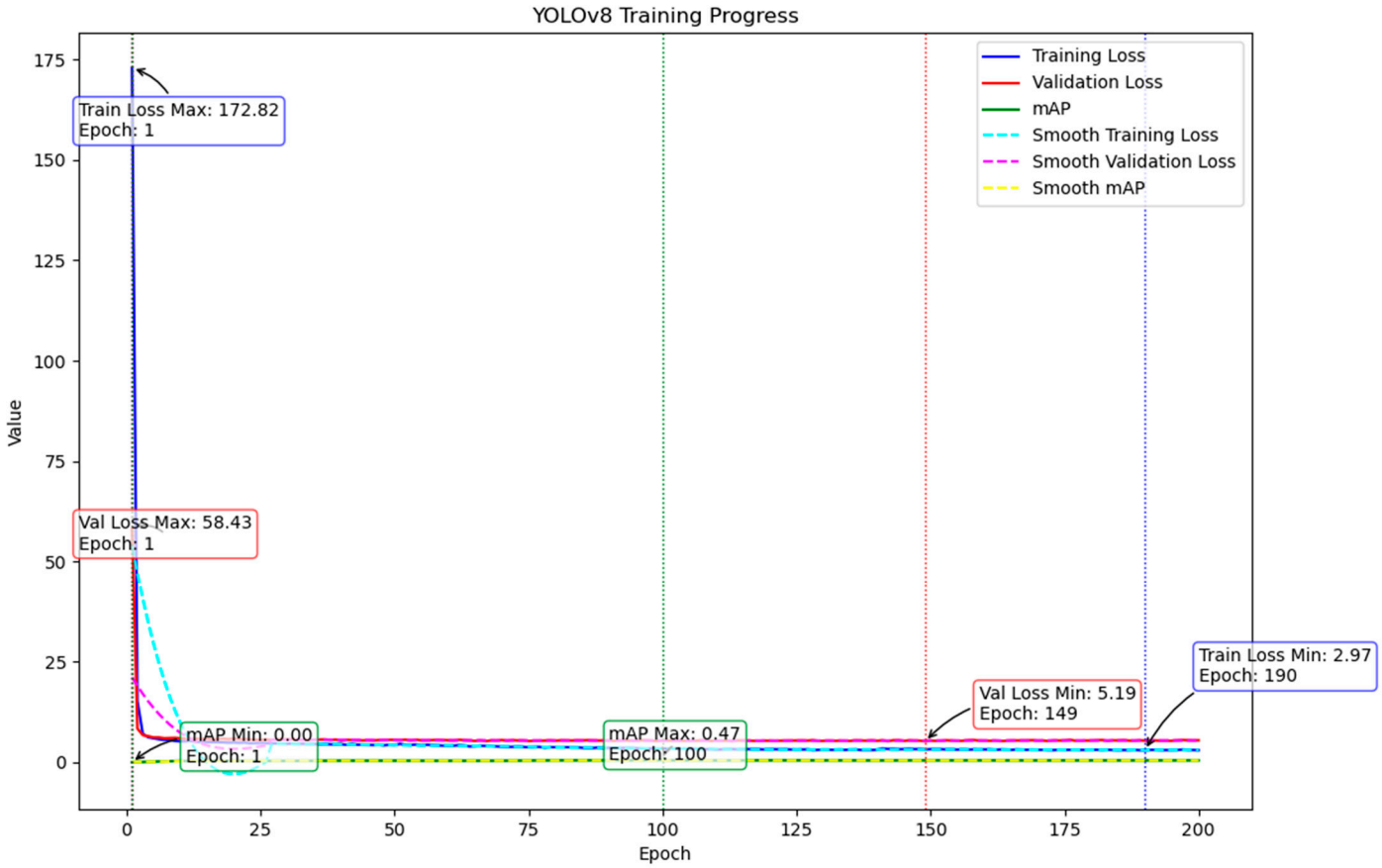


Figure 20. Loss value statistics obtained during model training.

The training loss reaches its maximum value in the first epoch of training (172.82), indicating that the model's error is large in the initial stage. As training progresses, the training loss rapidly decreases, reaching a minimum value of 2.97 at the 190th epoch. This significant decrease in training loss indicates that the model gradually optimizes its parameters during the continuous learning and adjustment process, thereby reducing errors.

The change trend for validation loss is similar to that for training loss, reaching a maximum value of 58.43 in the first epoch. As training progresses, validation loss gradually decreases and reaches a minimum value of 5.19 in the 149th epoch. The downward trend of validation loss indicates that the model's performance on unseen data is also improving, indicating that it has good generalization ability.

The mAP is the smallest in the first epoch of training, amounting to 0.00, indicating that the initial model has poor performance in the detection task. As the training progressed, the mAP gradually increased and reached a maximum value of 0.47 in the 100th epoch. Although the mAP fluctuates slightly in the later stage, it remains at a high level overall, indicating that the accuracy of the model in various types of damage detection continues to improve.

This study's detailed analysis of the loss value during the training process reveals a continuous optimization of the model's performance on the training and validation sets, significant decreases in the training and validation losses, and a significant increase in the mAP. We used the 100th-, 149th-, and 190th-epoch models, as well as the final, 200th-epoch model with outstanding numerical performance, as models for in-depth research and testing to further verify the model's performance. We will conduct a more comprehensive evaluation of the model's stability and generalization ability using these representative models to ensure its reliability and effectiveness in practical applications. This will provide more solid technical support for the efficient detection of damage to wooden structures in Fujian Tulou.

4.3. Comparative Analysis of Model Test Indicators

Based on the research results noted in the previous section, we selected models with outstanding numerical performance in the 100th, 149th, and 190th epochs, as well as models in the final (200th) epoch, for in-depth research and testing. If researchers wish to fully understand how well these models perform in detecting objects at different training stages (Figure 21), they need to examine key indicators like average precision, log-average miss rate, F1, recall, and precision.

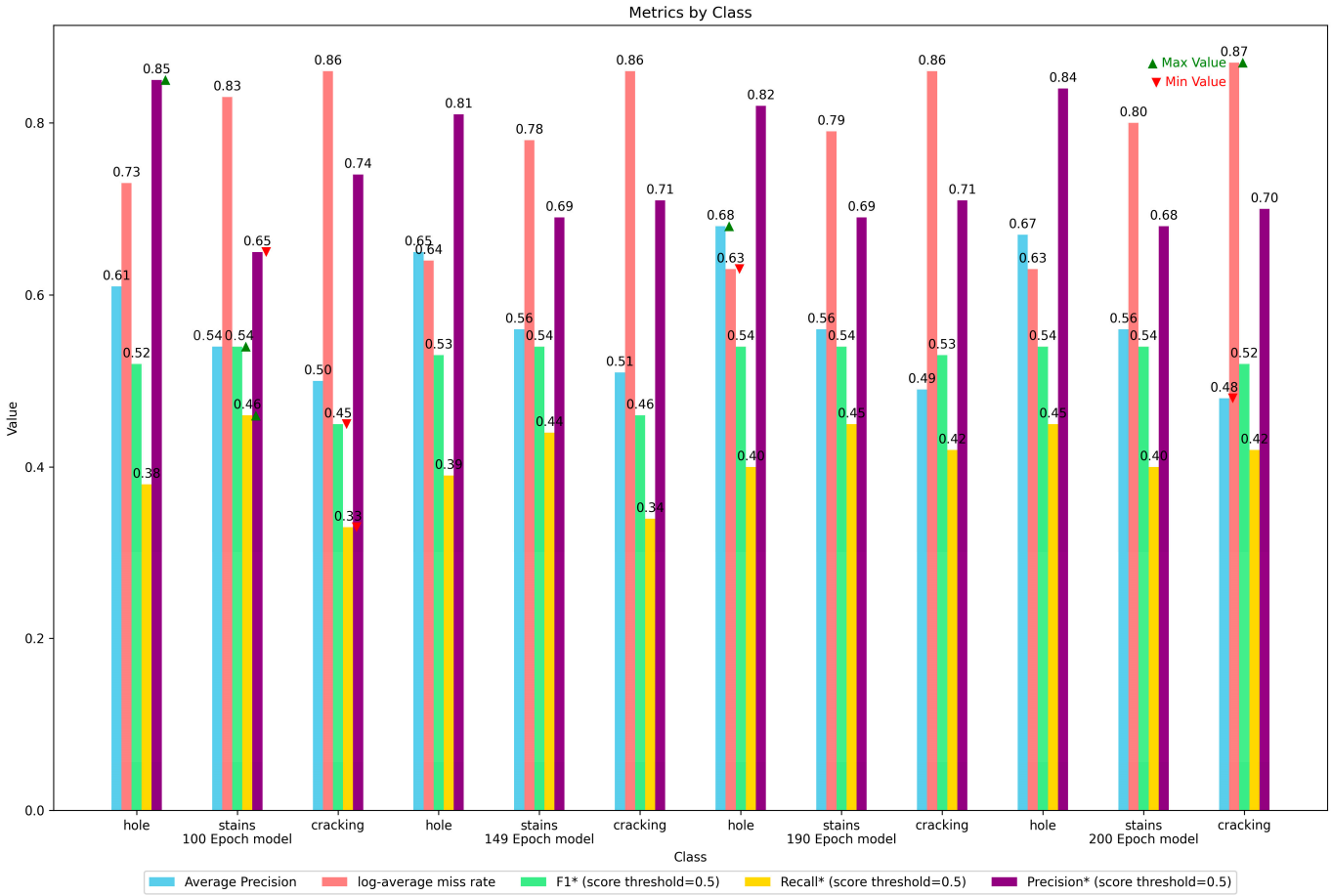


Figure 21. Model test results at different times (asterisk * in the figure indicates the median). In the figure, F1* indicates score threshold = 0.5; Recall* indicates score threshold = 0.5; Precision* indicates score threshold = 0.5.

(1) In the 100th epoch, the models' average precision for the hole label is 0.61, the log-average missed detection rate is 0.73, the F1 value is 0.52, the recall rate is 0.38, and the precision is 0.85. These results show that the model has some accuracy in hole detection, but it has a high missed detection rate and a low recall rate, which means that the model misses many hole samples. Although its precision is high, its insufficient F1 value and recall rate limit its overall performance. For the stains label, the average precision is 0.54; the log-average missed detection rate is 0.83; the F1 value and recall rate are 0.54 and 0.46, respectively; and the precision is 0.65, showing that the model performs well in stain detection, with a high missed detection rate. Although the recall rate and F1 value are relatively balanced, there is still room for improvement. The crack label has an average precision of 0.50, a log-average missed detection rate of 0.86, an F1* value of 0.45, a recall rate of 0.33, and a precision of 0.74. The model's performance in crack detection is poor, particularly due to its low recall rate and high missed detection rate. Despite the high accuracy, there is still room for improvement in the overall performance.

The confusion matrix of the 100th-epoch model demonstrates the model's classification performance for different labels (Figure 22). For the cracking label, the correct classification rate is 0.55, indicating that the model can correctly identify 55% of the cracking samples. Although this is a relatively high classification rate, 37% of the cracking samples are still misclassified as background, demonstrating that the model has obvious deficiencies in distinguishing between cracks and background. The model effectively distinguishes between cracking and holes, as evidenced by the misclassification of only 0.01 of the cracking samples as holes.

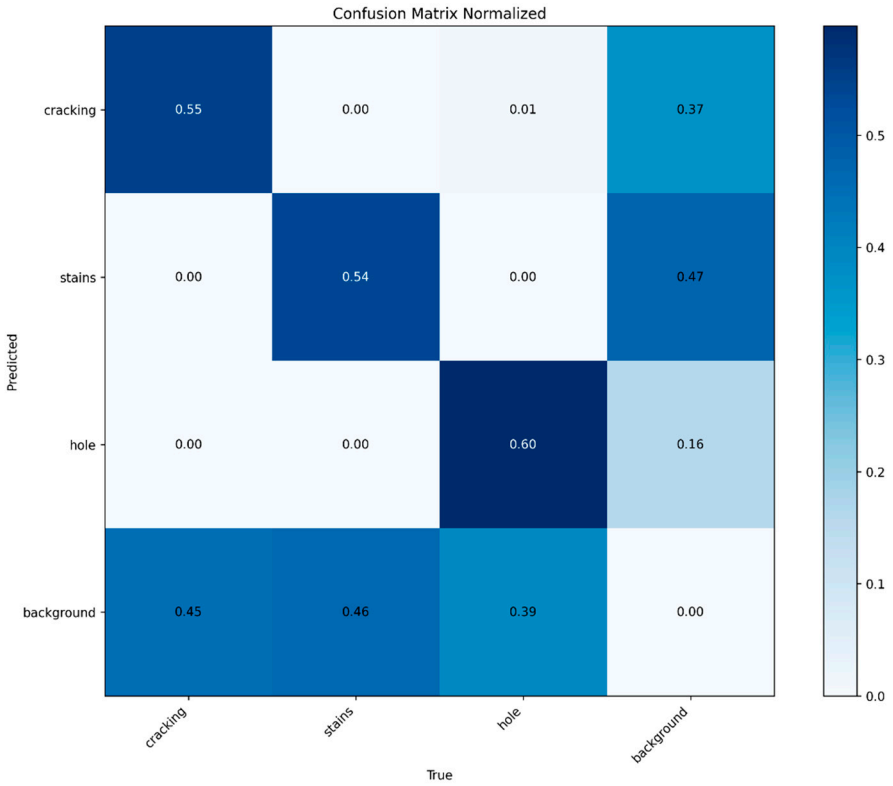


Figure 22. The confusion matrix is part of the 100th epoch.

(a) For the stains label, the correct classification rate is 0.54, which is comparable to the classification rate for the crack label, indicating that the model performs relatively stably in stain detection. However, the model misclassifies 47% of the stain samples as background, demonstrating significant errors in distinguishing stains from background. This high false-positive rate affects the model's overall detection performance.

(b) For the hole label, the correct classification rate is 0.60, showing that the model performs well in hole detection. Despite being a relatively low value, 16% of the hole samples were misclassified as background, demonstrating the model's limitations in distinguishing holes from background. The model demonstrated strong distinguishing ability by correctly classifying all the hole samples, not classifying any as other labels.

(c) The model misclassifies most of the background samples as other labels, misclassifying 45% as cracks, 46% as stains, and 39% as holes. This shows that the model has serious deficiencies in distinguishing background from other labels.

(2) The performance of the 149th-epoch model on the hole label showed improvement, with an average precision of 0.65, a log-average missed detection rate of 0.64, an F1 value of 0.53, a recall rate of 0.39, and a precision of 0.81, showing that the missed detection rate and recall rate of the model improved to a certain extent when detecting holes, and the model's overall performance has improved significantly. For the stains label, the average precision

is 0.56, the log-average missed detection rate is 0.78, the F1 value is 0.54, the recall rate is 0.44, and the precision is 0.69. The accuracy and recall rate of the model in stain detection improved, and the missed detection rate decreased, showing the progress of the model. The average precision for the cracking label is 0.51, the log-average missed detection rate is 0.86, the F1 value is 0.46, the recall rate is 0.34, and the precision is 0.71. Although the average precision and F1 value of cracking detection improved, the missed detection rate is still high and needs further optimization.

The confusion matrix of the 149th-epoch model shows the classification performance of the model for different labels (Figure 23). For the cracking label, the correct classification rate is 0.57, which is slightly higher than the value of 0.55 for the 100th-generation model. The proportion of samples misclassified as background has dropped from 37% to 33%, indicating that the model has improved in distinguishing cracking from background. At the same time, the proportion of samples misclassified as holes remains at a low level (0.01), indicating that the model's ability to distinguish in this regard is still strong.

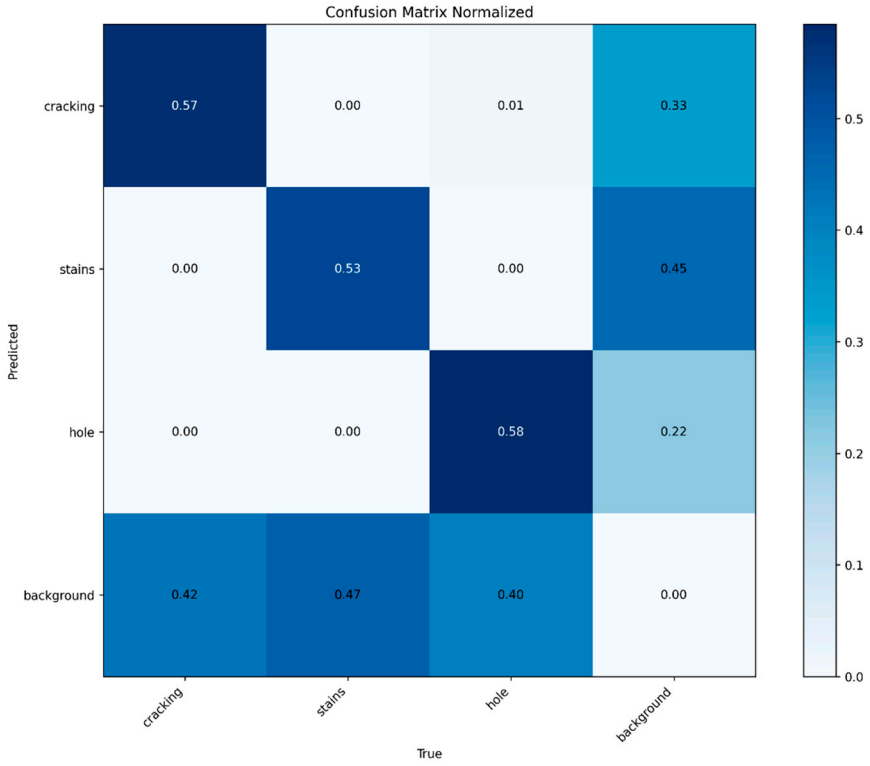


Figure 23. The confusion matrix for the 149th epoch.

(a) For the stains label, the correct classification rate is 0.53, which is slightly lower than the value of 0.54 for the 100th-epoch model. However, the proportion of images misclassified as background remains unchanged, amounting to 45%. Although the overall performance of the model on this label fluctuates slightly, the stability of the misclassification rate indicates that the model performs relatively consistently in this regard and needs further optimization to reduce misclassification.

(b) For the hole label, the correct classification rate is 0.58, which is slightly lower than the value of 0.60 for the 100th-epoch model. The proportion of images misclassified as background increased from 16% to 22%, which shows that the model's performance in distinguishing holes from background has decreased. This phenomenon may be due to changes in the distribution of the data during training or the adaptability of the model to specific features.

(c) The background misclassification rate is distributed among the various labels as follows: 42% of samples were classified as cracks, 47% of samples were classified as stains, and 40% of samples were classified as holes. Compared with the 100th-epoch

model, the proportion of background samples classified as stains and holes has increased slightly, showing that the model still has obvious deficiencies in distinguishing background from other labels.

(3) The 190th-epoch model performed remarkably better for each label. For the hole label, the average precision increased to 0.68, the log-average missed detection rate dropped to 0.63, the F1 value was 0.54, the recall rate was 0.40, and the precision was 0.82. The model performed well in hole detection, as both the missed detection rate and recall rate improved, and accuracy was maintained at a high level. The average precision of the stains label is 0.56, the log-average missed detection rate is 0.79, the F1 value is 0.54, the recall rate is 0.45, and the precision is 0.69. The accuracy and recall rate of the model in stain detection remain at a high level, and the missed detection rate improved slightly. The average precision of the cracking tag is 0.49, the log-average missed detection rate is 0.86, the F1 value is 0.53, the recall rate is 0.42, and the precision is 0.71. Although the average precision decreased slightly, the F1 value and recall rate were significantly improved, and the overall performance of the model in crack detection was significantly improved.

From the confusion matrix diagram of the 190th-epoch model, we can see the model's classification performance for different labels (Figure 24). For the cracking label, the correct classification rate is 0.58, which is slightly higher than the value of 0.57 for the 149th-epoch model. The proportion of images misclassified as background dropped from 33% to 36%, and while this shows a slight improvement in the model's ability to distinguish splits from background, there is still a significant misclassification problem. The model's performance in this regard improved compared to previous generations.

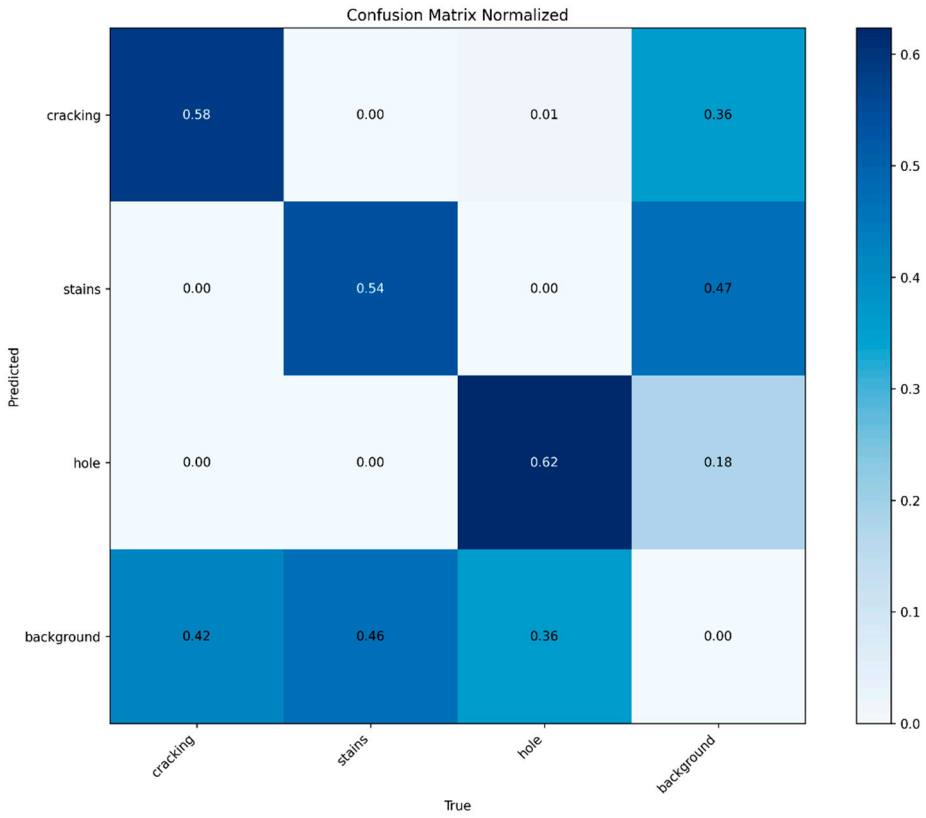


Figure 24. The confusion matrix for the 190th epoch.

(a) For the stains label, the correct classification rate is 0.54, a value consistent with the 149th-epoch model. The rate of misclassifying images as background is still 47%, indicating that the model has not significantly improved in terms of distinguishing stains from background. Although the model's overall performance for this label

is stable, the persistently high misclassification rate remains an issue that needs to be addressed.

(b) For the hole label, the correct classification rate is 0.62, which is significantly higher than the value of 0.58 corresponding to the 149th-epoch model. The proportion of images misclassified as background dropped from 22% to 18%, showing an improvement in the model's performance in distinguishing holes from background. This shows a significant improvement in the model's ability to detect holes.

(c) The distribution of the background misclassification rates for different labels is as follows: 42%, 46%, and 36% of samples were classified as cracks, stains, and holes, respectively. Compared with the 149th-generation model, the proportion of background samples classified as holes has decreased, showing that the model has improved in terms of handling the distinction between background and holes.

(4) The 200th-epoch model performed slightly better than the 190th-epoch model regarding the hole label, with an average precision of 0.67, a log-average false-positive rate of 0.63, an F1 value of 0.54, a recall rate of 0.45, and a precision of 0.84. Although the average precision and precision slightly decreased, the improvement in recall ensured that the F1 value remained unchanged, and the model still performed well regarding hole detection. The average precision of the stains label is 0.56, the log-average false-positive rate is 0.80, the F1 value is 0.54, the recall rate is 0.40, and the precision is 0.68. The model performed stably for stain detection, but the recall rate decreased slightly, affecting the overall performance. The average precision of the cracking label is 0.48, the log-average false-positive rate is 0.87, the F1 value is 0.52, the recall rate is 0.42, and the precision is 0.70. Although the F1* value and recall rate remained stable, the average precision and log-average false-positive rate deteriorated.

The confusion matrix of the 200th-epoch model shows the classification performance of the model across different labels (Figure 25). For the cracking label, the correct classification rate is 0.59, which is slightly higher than the value of 0.58 for the 190th-epoch model. The proportion of images misclassified as background remained 36%, the same as the value for the 190th-generation model. This shows that the model's performance in distinguishing cracks from the background remains stable and improved.

(a) For the stains label, the correct classification rate is 0.54, which is consistent with the previous epochs. The proportion of images misclassified as background slightly decreased to 46%. Although the overall performance of the model with respect to this label is relatively stable, the continued high misclassification rate is still an issue that needs to be addressed.

(b) For the hole label, the correct classification rate is 0.61, which is slightly lower than the value of 0.62 yielded in the 190th-epoch model. The proportion of images misclassified as background remained at 18%—the same as that for the 190th-generation model. These results show that the model's performance in distinguishing holes from the background remained stable.

(c) The background misclassification rate is distributed among the labels as follows: 41% of samples were classified as cracks, 46% were classified as stains, and 38% were classified as holes. Compared to the 190th-epoch model, the proportion of background samples classified as cracks and stains slightly decreased, while the proportion classified as holes increased, indicating that the model has fluctuated in distinguishing between the background from other labels.

In summary, combined with the indicator analysis and confusion matrix analysis of the 100th-, 149th-, 190th-, and 200th-epoch models, the following conclusions can be drawn: the 100th-generation model has a certain foundation in initial detection, but it had high misclassification and missed detection rates; the 149th-generation model exhibited an improvement compared to the 100th-epoch model, especially in cracking and hole detection, but the problem of misclassifying images as background was still serious; the 190th-epoch model exhibited a significant overall improvement over the performance of the first-generation

model, especially in terms of hole detection and crack detection, yielding the best performance, and the proportion of images misclassified as background was significantly reduced; and although the 200th-epoch model showed improvements in some aspects, its overall performance fluctuated, and it failed to comprehensively surpass the 190th-epoch model. Therefore, through this study, it has been shown that the 190th-epoch model performs well in various key indicators, especially in hole and crack detection, yielding the best overall performance, showing significant improvement and stable performance.

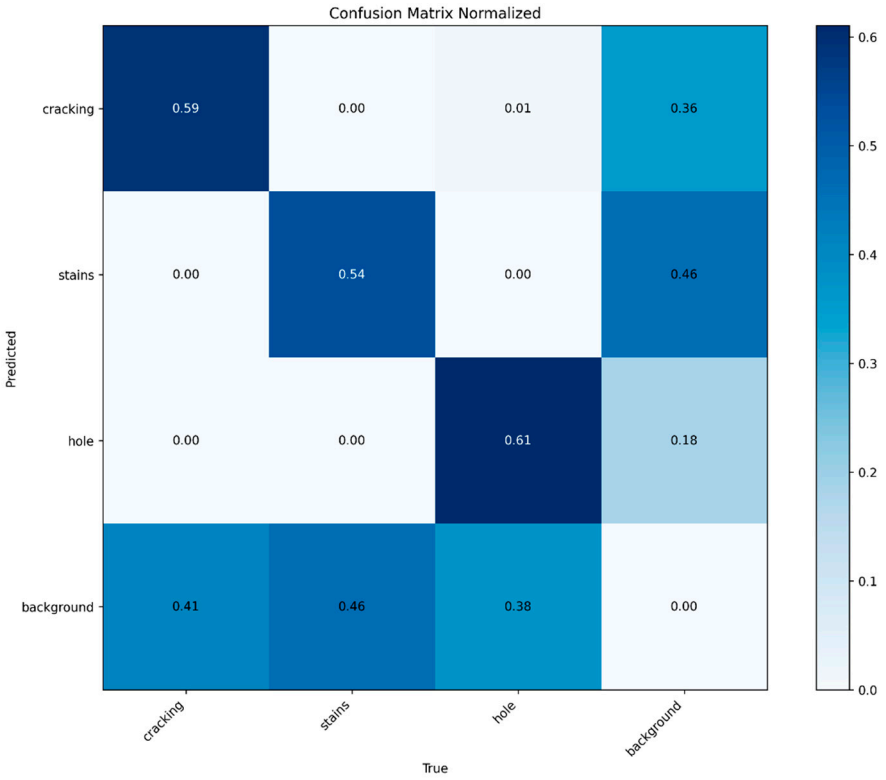


Figure 25. The confusion matrix for the 200th epoch.

4.4. Comparative Analysis of the Detection Results

To further verify the model's performance, this study selected three representative damage-type pictures and used the model from the 100th, 149th, 190th, and 200th epochs for detection tests. Figure 26 displays the results. Each picture represents a major damage type: holes, stains, and cracks. By comparing the detection results of different models in these pictures, this study can provide a deeper understanding of each model's performance in practical applications.

It should be noted that there may be differences between image tests and indicator tests. Indicator tests are usually performed on a specific validation set, while the real images used in actual tests may contain more variability and complexity. The noise, illumination changes, and non-standardized features in these real scenes may be more akin to the situation in actual applications, resulting in different performances of the model for the real test and for the standardized validation set. Therefore, in the previous studies, the 190th-epoch model performed best in the indicator test, but the results in the image test were not necessarily consistent with those in the indicator test. This section of the experiment will study this possible situation.

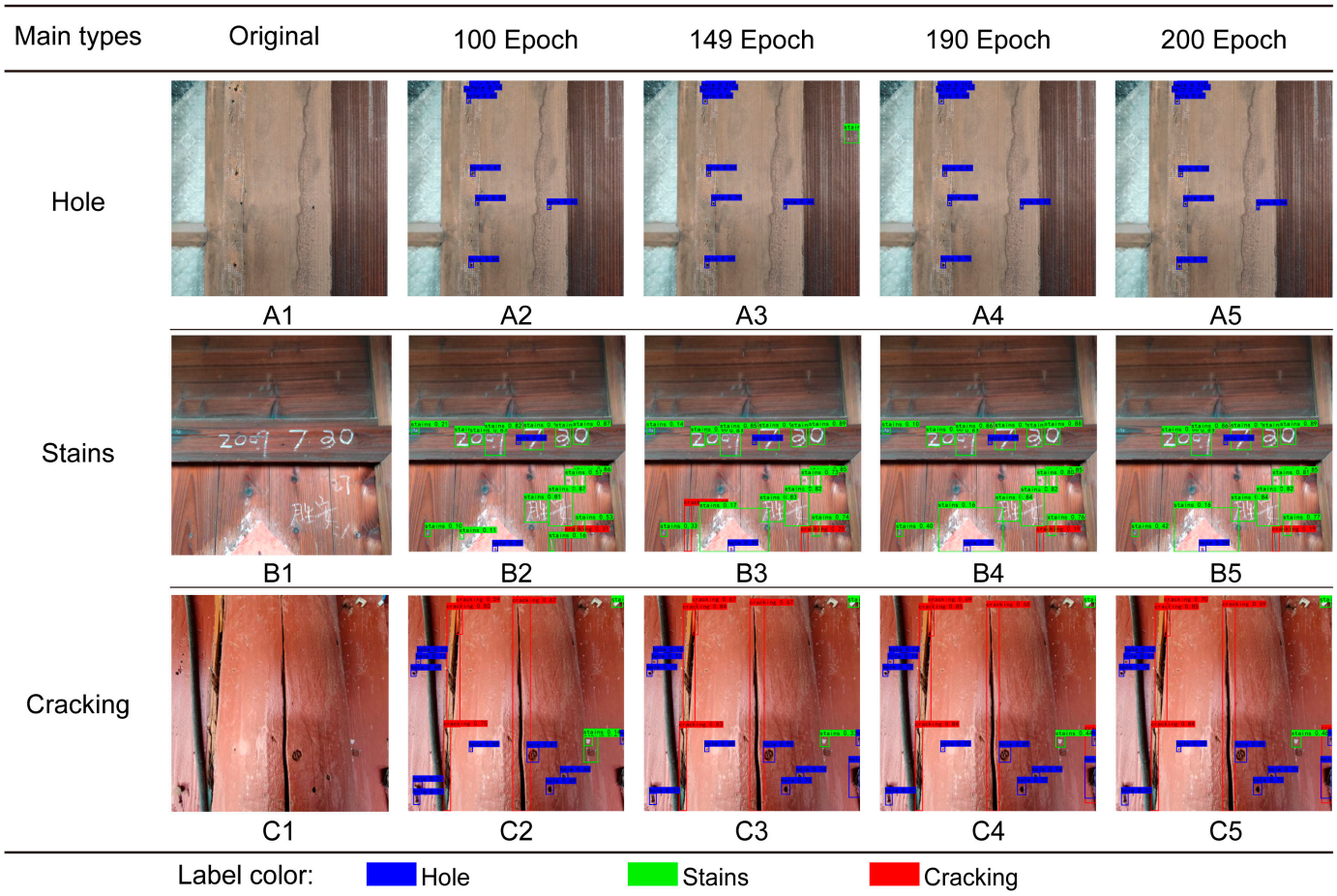


Figure 26. The results regarding the model's detection of wood structure damage at different epochs are presented.

In terms of detecting holes, the original picture, A1, shows multiple holes on the surface of a wooden structure. The 100th-epoch (A2) model successfully detected multiple holes, but there was some other damage that it failed to identify. The model of the 149th epoch (A3) performed best, detecting all major holes without significant false detections. The model of the 190th epoch (A4) and the model of the 200th epoch (A5) performed similarly in hole detection, but their performance was slightly lower than that of the model of the 149th epoch. The detection of individual damage types is prone to a phenomenon known as missed detection. Thus, the results show that the 149th-epoch model has the best performance in detecting holes and can provide higher detection accuracy and stability.

For stain detection, the original image, B1, shows that there are multiple stains (artificial graffiti) on the surface of the wooden structure. The model for the 100th epoch (B2) was able to detect most of the stains, but there was a certain amount of false detection. In stain detection, the model of the 149th epoch (B3) performed well, accurately identifying most of the stains while reducing the number of false detections. The models of the 190th (B4) and 200th epochs (B5) also performed well in stain detection but were slightly inferior to the model of the 149th epoch in this regard. Overall, the 149th-epoch model performed the most stably and accurately in stain detection.

In terms of crack detection, the original image, C1, shows obvious cracks on the surface of the wooden structure. The 100th-epoch model (C2) was able to identify most of the cracks, but there was a certain degree of missed detection. In crack detection, the 149th-epoch model (C3) performed best, accurately identifying all major cracks with fewer false positives. The 190th-epoch model (C4) and the 200th-epoch model (C5) performed slightly worse than the 149th-epoch model in crack detection, with false positives. This further shows that the 149th-epoch model has the best performance in crack detection.

A comprehensive analysis of the above detection results reveals that the model of the 149th epoch performed well in the detection of three types of damage, namely, holes, stains, and cracks, showing high detection accuracy and stability. In contrast, the 100th-epoch model performed slightly worse in all types of damage detection, with some missed detections and false detections. The 190th- and 200th-epoch models were similar in terms of detection performance but yielded slightly lower values in stability and accuracy than the 149th-epoch models.

After a thorough analysis of the detection results for the three main types of damage images, we can conclude that the model of the 149th epoch performs best in practical applications and can accurately and stably detect damage types such as holes, stains, and cracks in wooden structures. The results show that the model trained for 149 epochs achieved the best performance in various damage detection tasks, providing reliable technical support for the efficient detection of damage to wooden structures in Fujian Tulou. This model can be used to continuously optimize and improve the detection effect through further research and application.

4.5. Analysis of Model Feature Layers

In order to deeply understand the working mechanism and detection performance of the model, we conducted a detailed analysis of the model's feature layers. Feature layers are different levels of features extracted by a model while processing an input image. By observing the feature layers, we can understand how our model gradually identifies and processes different types of wood structure damage. Figure 27 illustrates the response of the input image at various feature layers, specifically the cv2 and cv3 layers, along with the outcomes of feature fusion and the final target detection results. Feature layer analysis includes the observation of small-scale, medium-scale, and large-scale feature maps, which correspond to feature extraction from the cv2 and cv3 layers, respectively.

The input image is a 512 $\times$ 512-pixel image of a wooden structure surface, which contains obvious damage types such as stains, cracks, and holes. To detect objects of different sizes, the YOLOv8 model uses its detection head to extract multi-scale features. The detection head is divided into a regression branch and a classification branch, which use different loss functions: the regression branch uses Distribution Focal Loss and CIoU Loss, while the classification branch uses BCE Loss.

In the small-scale feature map, the model uses the cv2[0] and cv3[0] layers to extract detailed features from the image. The regression branch of the cv2[0] layer primarily predicts the target's location and size. The feature map shows that the model responds strongly to small and dense targets. For example, in the cv2[0] feature map, the model responds strongly to stains, which are usually small and densely distributed. The small-scale feature map can clearly identify these areas' details. The cv3[0] layer serves as a classification branch, primarily utilizing BCE Loss to predict the target's category. The feature map shows that the model has a high response to stains, indicating that the model can capture subtle and dense stain features in this layer.

The cv2[1] and cv3[1] layers extract the mesoscale feature maps, which correspond to medium-sized object detection. The cv2[1] feature map shows that the model has a strong response to medium-sized stains and cracked areas, indicating that the model can capture more obvious damage features at the mesoscale, especially those that are neither too small nor too large. In the cv3[1] feature map, the model's response to cracking is particularly prominent, showing that the model can effectively identify medium-scale crack features, which are usually long and of moderate width.

The cv2[2] and cv3[2] layers extract large-scale feature maps, primarily for detecting larger objects. In the cv2[2] feature map, the response is more scattered, but it can capture some large-area damage features in the image, especially holes and cracks. The cv3[2] feature map has no obvious thermal response. These large-scale features usually require larger objects in the image in order to function.

The feature fusion graph depicts the model's comprehensive responses at different feature layers. By fusing the features of the cv2 and cv3 layers, the model can more comprehensively identify various types of damage in the image. The highlights in the feature fusion graph include the high-response areas in the first two layers of the feature graphs, but they also integrate features at different levels, providing more comprehensive and accurate damage detection results.

The final object detection result shows the model's detection and annotation of the input image. The model successfully identified and annotated all the main types of damage in the image. These results show that through the gradual extraction and fusion of feature layers, the model can effectively identify and locate different types of wood structure damage, providing reliable technical support for practical applications.

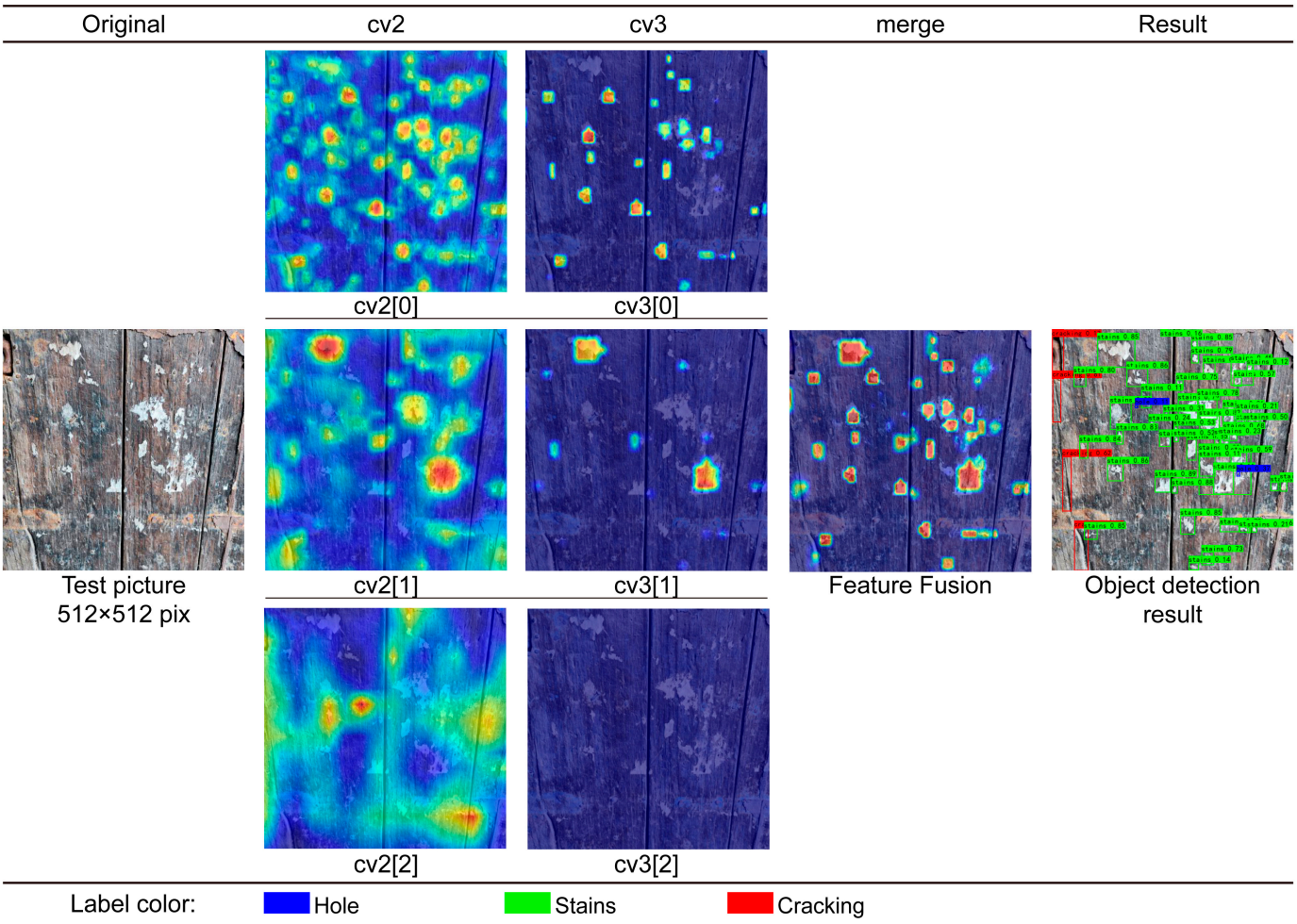


Figure 27. Model feature map testing.

4.6. Model Application

To test the model's detection performance in practical applications, we used an uncropped on-site photo. The original dimensions of the image are 3000 × 4000 pixels, and it contains various types of damage to the surface of the wooden structure. By analyzing the responses of different feature layers, this study can provide insight into how the model handles and identifies damage to wooden structures at different levels (Figure 28).

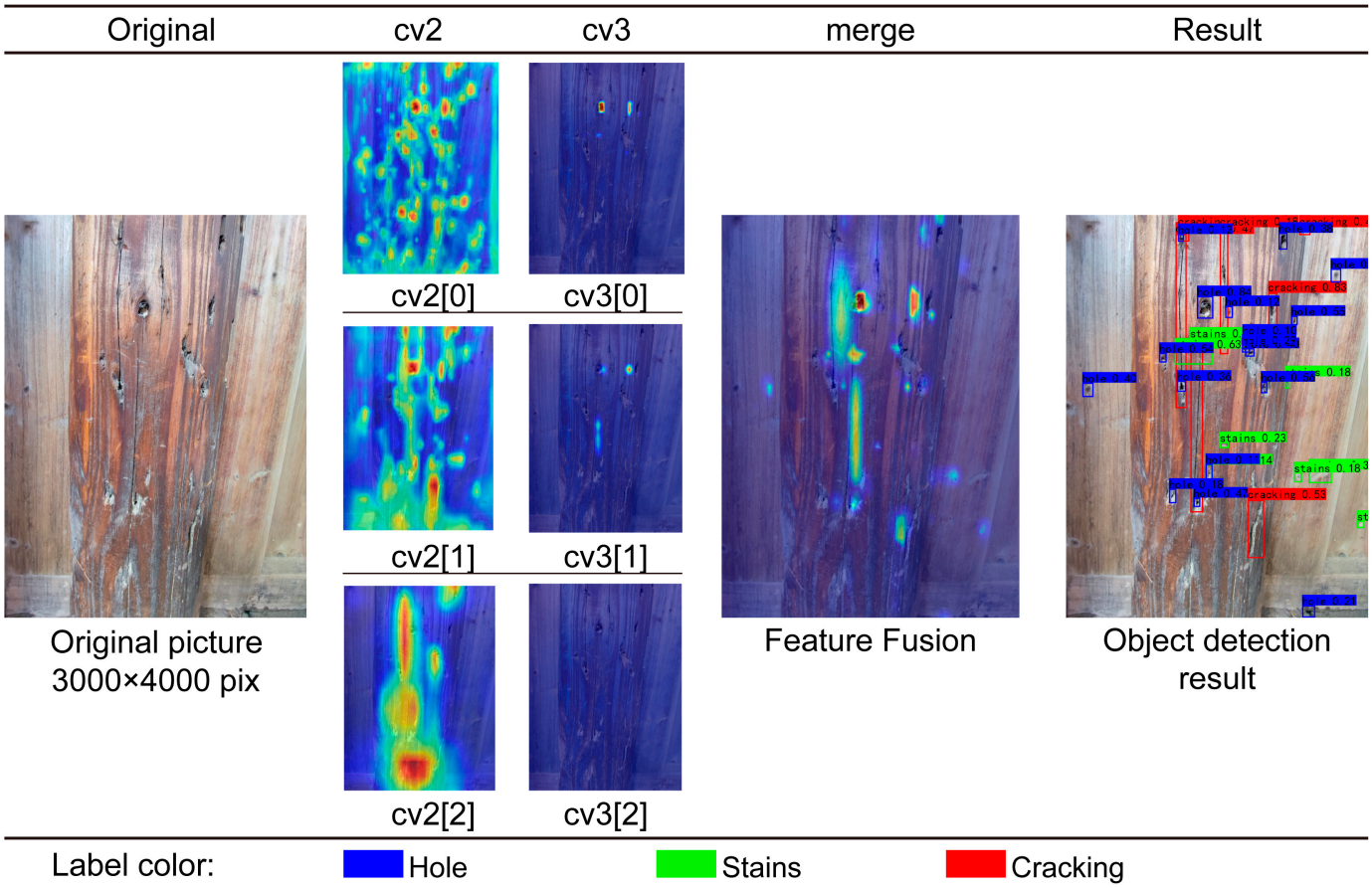


Figure 28. Characterization of field-test models.

In the small-scale feature map, cv2[0] and cv3[0] show the responses of the regression branch and the classification branch, respectively. The main purpose of cv2[0] is to predict the location and size of the target. The concentration of bright spots in the hole area of the feature map suggests that the model can effectively capture small and dense damage features. We used cv3[0] to predict the target's category. The feature map's high response to the hole and crack areas demonstrates that the model can effectively identify subtle damage of the wood structure on a small scale.

The mesoscale feature map includes the cv2[1] and cv3[1] layers, which are mainly used to detect medium-sized objects. The cv2[1] feature map concentrates bright spots in medium-sized stains and cracks, demonstrating the model's accuracy in locating and identifying these types of damage at the mesoscale. The cv3[1] layer's response demonstrates that the model has a strong recognition ability for medium-sized cracks, which are usually long and of moderate width. This layer allows the model to capture their main features.

Large-scale feature maps were extracted from cv2[2] and cv3[2] and used to detect larger objects. The widely distributed bright spots in the cv2[2] feature map demonstrate the model's ability to identify large-area damage features, particularly large holes and cracks, even at a low resolution. The cv3[2] layer has almost no thermal response, indicating that the detection target is not suitable for large-scale feature layers.

The feature fusion map combines the responses of each feature layer and provides a comprehensive damage detection result by integrating the features of the cv2 and cv3 layers. After the model integrates multi-scale features, the bright spot area represents the high-response area, which can more accurately identify and locate various types of damage in an image.

The final object detection results show the model's detection annotation of the input image, successfully identifying and annotating all major damage types in the image. This

multi-scale feature fusion method allows the model to efficiently detect various types of damage to wooden structures in high-resolution site photos, and this analysis verifies the model's ability to do so. The analysis of these feature layers shows the model's responses at different levels, further confirming its robustness and accuracy.

Next, we tested the model's detection performance in an actual application. We selected the KuiJu Lou in Fujian Tulou for field testing. We designed the test to assess the model's ability to detect damage in real, complex environments. Figure 29 shows the results of the field test. We can analyze the model's performance in detail by testing different wooden columns. This study selected six representative wooden columns for testing, labeled A, B, C, D, E, and F. The original image and test result diagrams for each column show the model's detection ability in the actual application.

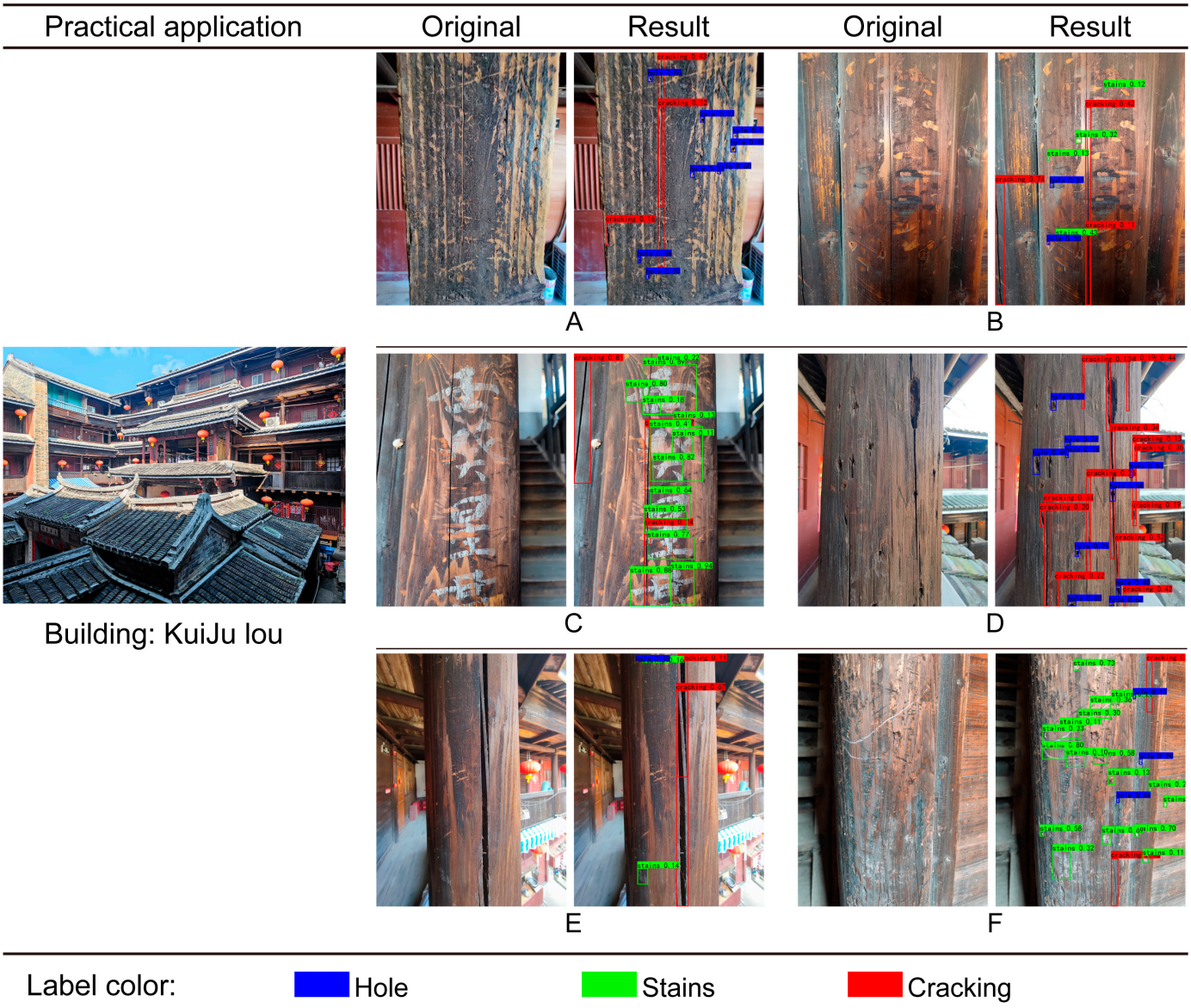


Figure 29. Results of the field tests.

The test results in column A show that the model successfully identified multiple holes (marked in blue) and cracks (marked in red). From the original image to the result image, it can be seen that the model accurately located these types of damage, especially in the complex texture background, and it could still effectively detect small holes and cracks.

The detection results in column B also demonstrate the model's strong performance. In the original image, there are many types of damage on the surface of the wooden column, as well as shadows covering it. As shown in the result image, the model accurately identified these damage types, particularly the stains and cracks that remain unaffected by shadows, demonstrating the model's reliability in handling intricate damage features.

The inspection results for column C further confirm the model's stability. The result image clearly marks the large stains and elongated cracks from the original image, demonstrating the model's high accuracy in identifying these types of damage. Among the types of damage, the model performs well in detecting cracks and can still identify partially hidden cracks even when covered by large stains.

The inspection results in column D demonstrate the model's comprehensive detection capabilities for multiple types of damage and its ability to resist interference from irrelevant elements. The original image clearly shows holes and cracks on the wooden column's surface, as well as a specific area in the background. The model accurately marked these damaged locations in the result image. Especially in terms of hole detection, the model can accurately locate multiple densely distributed holes, showing its advantage in detail processing.

The test results in column E show that the model is also good at detecting cracks and stains on wooden columns. The result map clearly marks all major damaged areas, especially long cracks. The model has efficient detection capabilities, demonstrating its wide applicability in detecting different types of damage.

The detection results shown in column F demonstrate the model's ability to detect a variety of small stains. The original image contains complex textures and multiple types of damage. The model accurately marked all major types of damage, as shown in the result image, demonstrating its efficiency and stability in dealing with complex scenes.

The on-site measurement of the wooden columns of the KuiJu Lou building validated the damage detection capability of the model in a complex real-world environment. The model can accurately identify and annotate various types of damage, such as holes, stains, and cracks in a wooden structure, showing its reliability and efficiency in practical applications. These test results further prove the model's superior performance in detecting damage to wooden structures and provide strong technical support for Fujian Tulou protection and restoration.

5. Discussion: The Application of Historic District Preservation Measures

As earth-and-wood structures, Tulou are susceptible to damage from natural causes, such as rain and typhoons, which can easily damage these buildings (Figure 30). Furthermore, the damp environment under the local natural conditions can also damage these buildings. Therefore, in the context of modern life, Tulous, whose facilities cannot meet modern living standards, were often abandoned and left idle before being included in the list of World Heritage Sites [89], creating a vicious cycle due to a lack of maintenance or insufficient funding.

Since the beginning of the twenty-first century, researchers have initiated regional protection measures and conducted sustainable research on Tulous and their surrounding areas by analyzing past research [90]. Relevant studies include those involving strengthening the modernization and livability of Tulous through sustainable design [91], exploring preservation status and destruction mechanisms [92], and evaluating the architectural environment [93]. These efforts have deepened the study of Tulous and provided ideas for their further development. However, these studies have struggled to improve the overall disaster prevention, protection planning, and maintenance difficulties relating to Tulous and their surrounding areas under limited repair and protection funds, constituting a significant yet overlooked aspect of Tulou preservation.

Only 46 of the Fujian Tulous are currently on the World Heritage List. Many more represent provincial, municipal, and county-level cultural heritage, with many buildings not receiving enough attention and financial support. Studies on Tulous in other regions,

where most Tulou maintenance funds are tight, corroborate this fact [94]. The maintenance of earth-and-wood structures is a challenging task, especially since these unique structures need to accommodate regular tourist visits and modern life for local residents with limited funds. This reality has led to the current crude and non-compliant maintenance and reinforcement measures applied to Tolou (Figures 31–33).



Figure 30. Chaotic repair site.

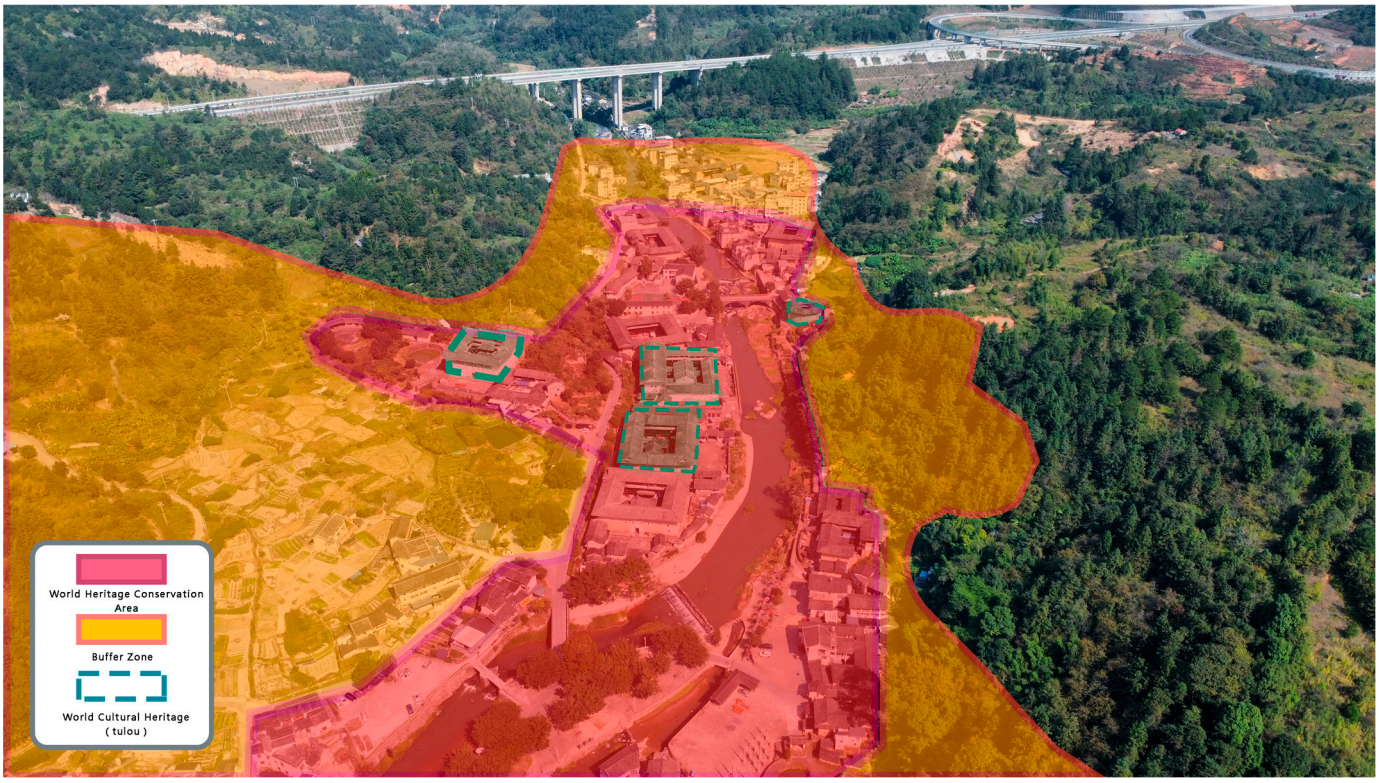


Figure 31. Schematic diagram of Tulou layout in Sampling Area 1.



Figure 32. Schematic diagram of Tulou layout in Sampling Area 2 (bird's eye view layout).

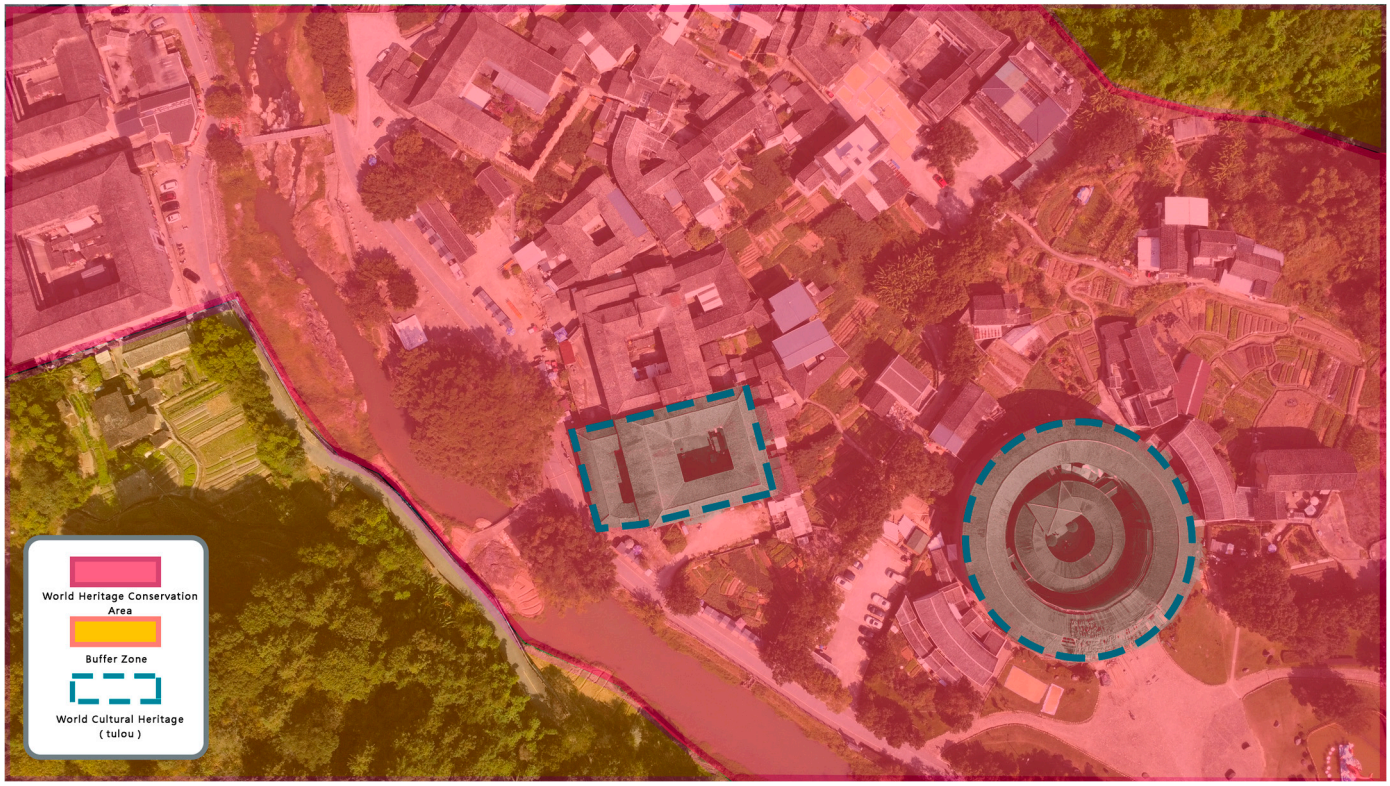


Figure 33. Schematic diagram of Tulou layout in Sampling Area 2 (floor plan).

Therefore, the local application of YOLO visual technology can prioritize the maintenance and care of severely damaged and heavily visited Tulous, thereby allocating limited

maintenance funds to these buildings and enhancing their sustainability. The dense aggregation of local wooden buildings in areas with Tulous also allows YOLO technology to be popularized locally at a low cost to prevent possible accidents caused by issues concerning wooden materials (Figures 34 and 35).



Figure 34. Complex wooden building area, showing possible disaster spread routes (overall).

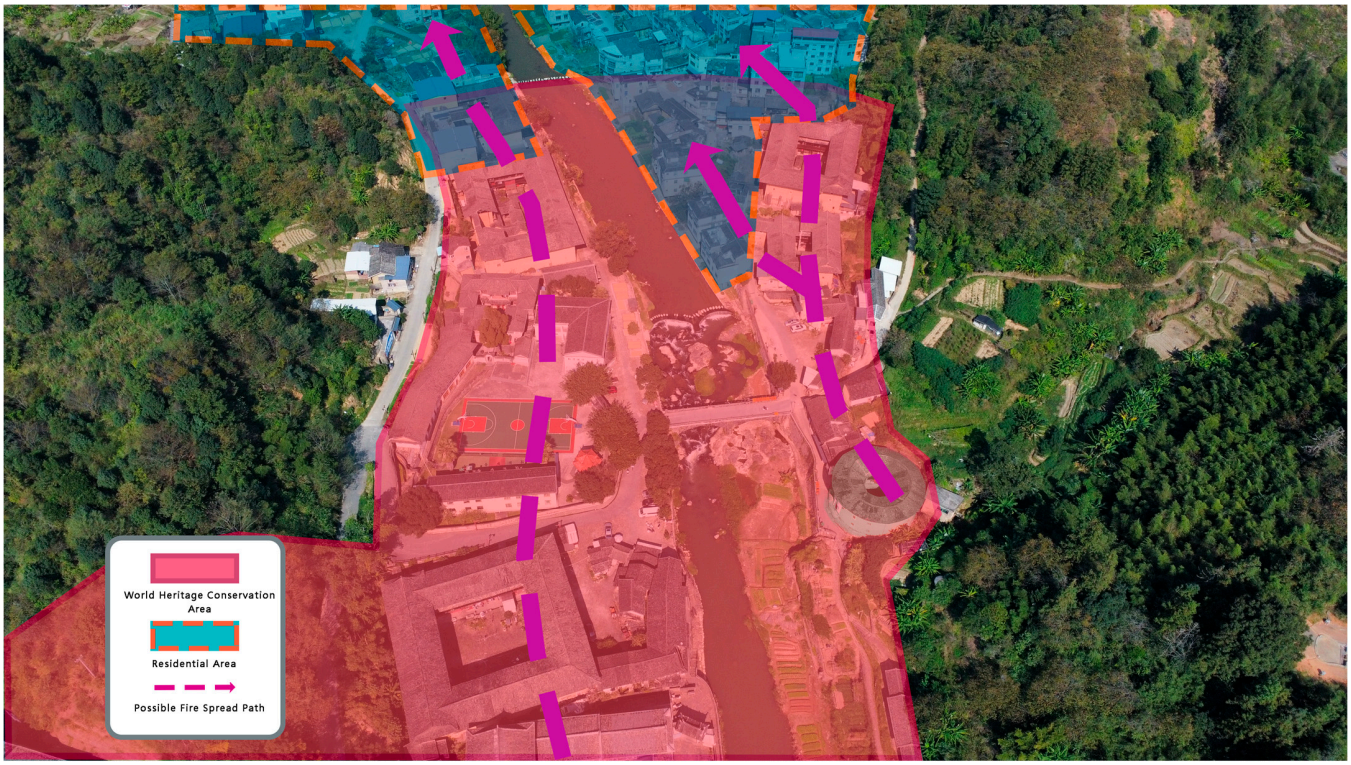


Figure 35. Complex wooden building area, showing possible disaster spread routes (partial).

6. Conclusions

Fujian Tulous, with their unique architectural features and historical value, have become a significant research subject in the context of global cultural heritage protection. Automatically detecting and analyzing damage to their wooden structures remain crucial tasks, especially when facing the challenges posed by rapidly changing natural materials and environments. This study explored the advantages of using computer vision and object detection technology to address this issue. By employing the YOLOv8 model and a diverse dataset consisting of over 1000 images, with 300 optimized for detailed preprocessing and annotation, we proposed and optimized a method of automatically detecting three types of damage in Fujian Tulous' wooden structures: cracking, stains, and holes. Through multiple experiments and adjustments, we gradually improved the model's detection performance and verified its effectiveness and reliability in practical applications.

6.1. Research Discoveries

The main conclusions and achievements of this study are as follows: (1) The YOLOv8 model was optimized through multiple experiments to allow it to perform well in wooden structure damage detection. By removing samples with complex backgrounds, improving label quality, and adjusting hyperparameters, the model's detection accuracy and stability were significantly improved. In the final experiment, the model achieved an overall mAP of 57.00% and was able to capture almost all damage points in the field test, meeting the needs of testing work. (2) Feature layer analysis demonstrated the model's powerful ability to process and identify damage to wooden structures at different scales. In uncropped high-resolution on-site photos, the model accurately identified and annotated various types of damage on wooden surfaces, verifying its effectiveness in practical applications. (3) In the field test conducted at KuiJu Lou in Fujian, the model performed well in complex environments and reliably detected damage types such as holes, stains, and cracks in wooden structures, confirming its efficiency and stability in practical applications and providing technical support for Fujian Tulou protection and restoration.

This research has three main advantages: (1) It enhances the efficiency and accuracy of cultural heritage protection by providing a method for the automated and intelligent detection of wooden structure damage, reducing errors, omissions, and workload associated with manual inspection. (2) It expands the application of computer vision in cultural heritage protection by verifying its feasibility and practicality through detailed experiments and analysis, providing valuable references and inspiration for future research. (3) It provides valuable insight and data support for further model optimization and adjustment through in-depth experimental analysis of the model's working principles and performance in detecting and identifying wooden structure damage.

6.2. Limitations and Future Work

Despite the remarkable results obtained, this study has some limitations:

(1) The model's accuracy in detecting cracks requires improvement, particularly when faced with complex backgrounds or light damage, leading to missed and false detections.

(2) The model's detection performance varies depending on the type of damage, with better results for some types (e.g., holes) and relatively poor results for others (e.g., stains and cracks).

(3) The current experiments focused mainly on the specific scenario of Fujian Tulous, and the model's adaptability and generalization ability for other types of wooden structures have not been fully verified.

Future research can be developed in the following directions:

(1) Dataset expansion: Collect more diverse wood structure damage data, especially from different environments and backgrounds, to improve the model's generalization ability.

(2) Model optimization: Explore advanced model architectures and training methods, such as multi-task learning and transfer learning, to improve detection performance for different damage types.

(3) Algorithm innovation: Investigate model structures and algorithms more suitable for assessing the damage characteristics of wooden materials to enhance accuracy and robustness.

(4) Multimodal learning: Combine data from various sources, such as drone imagery and images obtained via laser scanning, hyperspectral imaging, or X-ray imaging, to achieve more comprehensive and accurate detection.

(5) Practical application and system deployment: Focus on the model's deployment and optimization in real-world scenarios, integrating it into mobile terminals, drone devices, or portable devices for on-site technical support and decision-making assistance.

(6) Cross-scenario application: Apply the model to other types of wooden structures to verify its adaptability and stability in different application scenarios and expand its scope of application.

Overall, in this study, we created a useful model for finding damage to wooden structures by conducting many experiments and making improvements. The results show that the YOLOv8-model-based detection method we developed has a lot of potential for protecting cultural items in World Heritage historical blocks. Despite some shortcomings, further research and improvement will broaden the model's application prospects in wood structure damage detection. Furthermore, in the process of monitoring and detecting damage to architectural heritage, one frequently encounters interference from noise or other uncertain factors, typically unrelated to the structural degradation itself. This situation may lead to inaccurate or misunderstood monitoring results, which in turn affects the reliability of damage assessment. Therefore, improving the robustness of intelligent algorithms to these uncertain interferences becomes a key challenge. Future work will focus on developing more efficient algorithms to distinguish between structural degradation signals and unrelated noise or interference to improve the accuracy and reliability of detection. This will not only improve current monitoring technology, but also provide more solid scientific and technological support for heritage protection. In addition, future research can also consider exploring the further application of machine learning and artificial intelligence technologies in this field, especially the potential advantages in data processing and real-time monitoring. We also hope that this method, when combined with the renewal and protection of historical blocks, will play a greater role in the future.

Author Contributions: Conceptualization, J.F., Y.C. and L.Z.; methodology, L.Z.; software, J.F. and L.Z.; validation, J.F. and L.Z.; formal analysis, J.F., Y.C. and L.Z.; investigation, J.F.; resources, J.F., Y.C. and L.Z.; data curation, J.F., Y.C. and L.Z.; writing—original draft preparation, J.F., Y.C. and L.Z.; writing—review and editing, J.F., Y.C. and L.Z.; visualization, J.F., Y.C. and L.Z.; supervision, J.F., Y.C. and L.Z.; project administration, J.F., Y.C. and L.Z.; funding acquisition, Y.C. and L.Z. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by Guangdong Provincial Department of Education's key scientific re-search platforms and projects for general universities in 2023: Guangdong, Hong Kong, and Macau Cultural Heritage Protection and Innovation Design Team (grant number: 2023WCXTD042).

Data Availability Statement: The datasets used and analyzed during the current study are available from the corresponding author on reasonable request.

Conflicts of Interest: The authors declare no conflicts of interest.

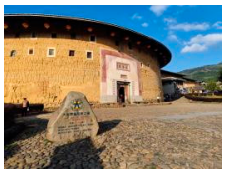
Appendix A. Machine Learning Environment

Machine learning environment: The operating system is Windows 11 (X64), the CUDA version is 11.5, the deep learning framework is PyTorch (1.13.0), and the graphics card and processor are a GeForce GTX 3070 (16 G) and an AMD Ryzen 9 5900HX (3.30 GHz), respectively.

Appendix B. Location Statistics of Image Collection during Fieldwork

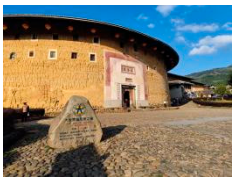
The following are details of the locations where photos were taken for this study.

Table A1. Statistics of photos collected in the first experiment.

Acquisition Time	Name and Location	Picture	Collection Position	Quantity
5 November 2023	Chaoyang Lou		Corridors	45
5 November 2023	Chengqi Lou		Corridors	797
5 November 2023	Fuxing Lou		Corridors	21
5 November 2023	Kuiju lou		Corridors	288
6 November 2023	Qingcheng Lou		Courtyards and corridors	92
6 November 2023	Yanxiang Lou		Courtyards and corridors	186
6 November 2023	Zhencheng Lou		Courtyards and corridors	77

Source: Author statistics.

Table A2. Statistics of photos collected in the second experiment.

Acquisition Time	Name and Location	Picture	Collection Position	Quantity
11 May 2024	Chaoyang Lou		Corridors	9
11 May 2024	Chengqi Lou		Corridors	80
11 May 2024	Fuxing Lou		Corridors	15
11 May 2024	Kuiju lou		Corridors	71
10 June 2024	Qingcheng Lou		Courtyards and corridors	73
10 June 2024	Yanxiang Lou		Courtyards and corridors	42
10 June 2024	Zhencheng Lou		Courtyards and corridors	10

Source: Author statistics.

References

1. UNESCO. Fujian Tulou. World Heritage Convention. 2008. Available online: https://whc.unesco.org/en/list/1113/ (accessed on 12 May 2024).
2. Hui, Z.; Shiyu, D. Impact of Ethnic Migration on the Form of Settlement: A Case Study of Tuiou of South Fujian and the Hakkas. *J. Landsc. Res.* **2015**, *7*, 75.
3. Huang, H. *Fujian's Tulou: A Treasure of Chinese Traditional Civilian Residence*; Springer Nature: Berlin/Heidelberg, Germany, 2019.
4. Zhang, D. The New Architectural Trend in China: The Heritage and Delopment of Traditional Culture. Ph.D. Dissertation, Miami University, Oxford, OH, USA, 2016. Available online: http://hdl.handle.net/2374.MIA/5988 (accessed on 12 May 2024).
5. Wang, S.-S.; Li, S.-Y.; Liao, S.-J. The Genes of Tulou: A Study on the Preservation and Sustainable Development of Tulou. *Sustainability* **2012**, *4*, 3377–3386. [CrossRef]
6. Lin, X.; Wu, Y. Architectural Spatial Characteristics of Fujian Tubao from the Perspective of Chinese Traditional Ethical Culture. *Buildings* **2023**, *13*, 2360. [CrossRef]
7. Shu, Y.; He, Y. Research on the historical and cultural value of and protection strategy for rammed earth watchtower houses in Chongqing, China. *Built Herit.* **2021**, *5*, 23. [CrossRef]
8. Zhou, Q. Research on Traditional Reinforcement Techniques for Rammed Earth Walls in China. *Int. J. Archit. Herit.* **2024**, 1–19. [CrossRef]
9. Elvin, M. *The Pattern of the Chinese Past: A Social and Economic Interpretation*; Stanford University Press: Redwood City, CA, USA, 1973.
10. Zhang, P.C.; Luo, K.; Liao, W.B. Study on the Material and the Structure of Earth Building in Fujian. *Adv. Mater. Res.* **2012**, *368*, 3567–3570. [CrossRef]
11. Porretta, P.; Pallottino, E.; Colafranceschi, E. Minnan and Hakka Tulou. Functional, typological and construction features of the rammed earth dwellings of Fujian (China). *Int. J. Archit. Herit.* **2022**, *16*, 899–922. [CrossRef]
12. Luo, Y.; Yin, B.; Peng, X.; Xu, Y.; Zhang, L. Wind-rain erosion of Fujian Tulou Hakka earth buildings. *Sustain. Cities Soc.* **2019**, *50*, 101666. [CrossRef]
13. Tan, C.; Ke, Y. Old houses in Hankou Concession: Collective memory and nostalgic space consumption in the context of modernization. *Herança* **2024**, *7*, *in press*.
14. Al Jaff, A.A.M.; Al Shabander, M.S.; Bala, H.A. Modernity and tradition in the context of Erbil old town. *Am. J. Civ. Eng. Archit.* **2017**, *5*, 217–224. [CrossRef]
15. Stubbs, M. Heritage-sustainability: Developing a methodology for the sustainable appraisal of the historic environment. *Plan. Pract. Res.* **2004**, *19*, 285–305. [CrossRef]
16. Nocca, F. The Role of Cultural Heritage in Sustainable Development: Multidimensional Indicators as Decision-Making Tool. *Sustainability* **2017**, *9*, 1882. [CrossRef]
17. Lovelady, A. Broadened notions of historic preservation and the role of neighborhood conservation districts. *Urban Lawyer* **2008**, *40*, 147.
18. Jokilehto, J. *A History of Architectural Conservation*; Routledge: London, UK, 2017. [CrossRef]
19. Bandarin, F.; Van Oers, R. *The Historic Urban Landscape: Managing Heritage in an Urban Century*; John Wiley & Sons: Hoboken, NJ, USA, 2012.
20. Tweed, C.; Sutherland, M. Built cultural heritage and sustainable urban development. *Landsc. Urban Plan.* **2007**, *83*, 62–69. [CrossRef]
21. Nocca, F.; Angrisano, M. The Multidimensional Evaluation of Cultural Heritage Regeneration Projects: A Proposal for Integrating Level(s) Tool—The Case Study of Villa Vannucchi in San Giorgio a Cremano (Italy). *Land* **2022**, *11*, 1568. [CrossRef]
22. Unesco. General Conference. In *Recommendation Concerning the Safeguarding and Contemporary Role of Historic Areas: Adopted by the General Conference at Its Nineteenth Session, Nairobi, 26 November 1976*; Unesco: Paris, France, 1976.
23. Lu, Y.; He, M.-E.; Liu, C. Tourism Competitiveness Evaluation Model of Urban Historical and Cultural Districts Based on Multi-Source Data and the AHP Method: A Case Study in Suzhou Ancient City. *Sustainability* **2023**, *15*, 16652. [CrossRef]
24. Owley, J. Cultural heritage conservation easements: Heritage protection with property law tools. *Land Use Policy* **2015**, *49*, 177–182. [CrossRef]
25. Maria, M.D.; Fiumi, L.; Mazzei, M.; V., B.O. A System for Monitoring the Environment of Historic Places Using Convolutional Neural Network Methodologies. *Heritage* **2021**, *4*, 1429–1446. [CrossRef]
26. Lin, H.; Huang, L.; Chen, Y.; Zheng, L.; Huang, M.; Chen, Y. Research on the Application of CGAN in the Design of Historic Building Facades in Urban Renewal—Taking Fujian Putian Historic Districts as an Example. *Buildings* **2023**, *13*, 1478. [CrossRef]
27. Li, Y.; Zhao, M.; Mao, J.; Chen, Y.; Zheng, L.; Yan, L. Detection and recognition of Chinese porcelain inlay images of traditional Lingnan architectural decoration based on YOLOv4 technology. *Herit. Sci.* **2024**, *12*, 137. [CrossRef]
28. Yan, L.; Chen, Y.; Zheng, L.; Zhang, Y. Application of computer vision technology in surface damage detection and analysis of shedthin tiles in China: A case study of the classical gardens of Suzhou. *Herit. Sci.* **2024**, *12*, 72. [CrossRef]
29. Zheng, L.; Chen, Y.; Yan, L.; Zhang, Y. Automatic Detection and Recognition Method of Chinese Clay Tiles Based on YOLOv4: A Case Study in Macau. *Int. J. Archit. Herit.* **2023**, *1*, 20. [CrossRef]
30. Li, Q.; Zheng, L.; Chen, Y.; Yan, L.; Li, Y.; Zhao, J. Non-destructive testing research on the surface damage faced by the Shanhaiguan Great Wall based on machine learning. *Front. Earth Sci.* **2023**, *11*, 1225585. [CrossRef]

31. Macek, D.; Heralová, R.S.; Hromada, E.; Střelcová, I.; Brožová, I.; Vitásek, S.; Pojar, J.; Bouška, R. Cost optimization for renovation and maintenance of cultural heritage objects. *IOP Conf. Ser. Earth Environ. Sci. IOP Publ.* **2019**, *290*, 012155. [CrossRef]
32. Zolkafli, U.K.; Zakaria, N.; Mohammad Mazlan, A.; Ali, A.S. Maintenance work for heritage buildings in Malaysia: Owners' perspectives. *Int. J. Build. Pathol. Adapt.* **2019**, *31*, 186–195. [CrossRef]
33. Li, Y.; Lu, Y.; Chen, J. A deep learning approach for real-time rebar counting on the construction site based on YOLOv3 detector. *Autom. Constr.* **2021**, *124*, 103602. [CrossRef]
34. Wei, G.; Wan, F.; Zhou, W.; Xu, C.; Ye, Z.; Liu, W.; Lei, G.; Xu, L. BFD-YOLO: A YOLOv7-Based Detection Method for Building Façade Defects. *Electronics* **2023**, *12*, 3612. [CrossRef]
35. Sarraf, A.; Azhdari, M.; Sarraf, S. A comprehensive review of deep learning architectures for computer vision applications. *Am. Sci. Res. J. Eng. Technol. Sci. (ASRJETS)* **2021**, *77*, 1–29.
36. Redmon, J.; Divvala, S.; Girshick, R.; Farhadi, A. You only look once: Unified, real-time object detection. In Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA, 27–30 June 2016; pp. 779–788. [CrossRef]
37. Chandana, R.K.; Ramachandra, A.C. Real time object detection system with YOLO and CNN models: A review. *arXiv* **2022**, arXiv:2208.00773. [CrossRef]
38. Jiang, P.; Ergu, D.; Liu, F.; Cai, Y.; Ma, B. A Review of Yolo algorithm developments. *Procedia Comput. Sci.* **2022**, *199*, 1066–1073. [CrossRef]
39. Narejo, S.; Pandey, B.; Esenarro Vargas, D.; Rodriguez, C.; Anjum, M.R. Weapon detection using YOLO V3 for smart surveillance system. *Math. Probl. Eng.* **2021**, *2021*, 1–9. [CrossRef]
40. Zhou, Y.; Wen, S.; Wang, D.; Meng, J.; Mu, J.; Irampaye, R. MobileYOLO: Real-Time Object Detection Algorithm in Autonomous Driving Scenarios. *Sensors* **2022**, *22*, 3349. [CrossRef]
41. Ünver, H.M.; Ayan, E. Skin Lesion Segmentation in Dermoscopic Images with Combination of YOLO and GrabCut Algorithm. *Diagnostics* **2019**, *9*, 72. [CrossRef]
42. Hussain, M. YOLO-v1 to YOLO-v8, the Rise of YOLO and Its Complementary Nature toward Digital Manufacturing and Industrial Defect Detection. *Machines* **2023**, *11*, 677. [CrossRef]
43. Aziz, L.; Salam, M.S.B.H.; Sheikh, U.U.; Ayub, S. Exploring deep learning-based architecture, strategies, applications and current trends in generic object detection: A comprehensive review. *IEEE Access* **2020**, *8*, 170461–170495. [CrossRef]
44. Siountri, K.; Anagnostopoulos, C.-N. The Classification of Cultural Heritage Buildings in Athens Using Deep Learning Techniques. *Heritage* **2023**, *6*, 3673–3705. [CrossRef]
45. Perez, H.; Tah, J.H.M.; Mosavi, A. Deep Learning for Detecting Building Defects Using Convolutional Neural Networks. *Sensors* **2019**, *19*, 3556. [CrossRef]
46. Doneus, M.; Verhoeven, G.; Fera, M.; Briese, C.; Kucera, M.; Neubauer, W. From deposit to point cloud–a study of low-cost computer vision approaches for the straightforward documentation of archaeological excavations. *Geoinform. FCE CTU* **2011**, *6*, 81–88. [CrossRef]
47. Altaweel, M.; Khelifi, A.; Li, Z.; Squitieri, A.; Basmaji, T.; Ghazal, M. Automated Archaeological Feature Detection Using Deep Learning on Optical UAV Imagery: Preliminary Results. *Remote Sens.* **2022**, *14*, 553. [CrossRef]
48. Jalandoni, A.; Zhang, Y.; Zaidi, N.A. On the use of Machine Learning methods in rock art research with application to automatic painted rock art identification. *J. Archaeol. Sci.* **2022**, *144*, 105629. [CrossRef]
49. Hosain, M.T.; Zaman, A.; Abir, M.R.; Akter, S.; Mursalin, S.; Khan, S.S. Synchronizing Object Detection: Applications, Advancements and Existing Challenges. *IEEE Access* **2024**, *12*, 54129–54167. [CrossRef]
50. Karimi, N.; Mishra, M.; Lourenço, P.B. Deep learning-based automated tile defect detection system for Portuguese cultural heritage buildings. *J. Cult. Herit.* **2024**, *68*, 86–98. [CrossRef]
51. Reinprecht, L. *Wood Deterioration, Protection and Maintenance*; John Wiley & Sons: Hoboken, NJ, USA, 2016.
52. Suzuki, M.; Miyauchi, T.; Isaji, S.; Hirabayashi, Y.; Naganawa, R. Decay detection of constructional softwoods using machine olfaction. *J. Wood Sci.* **2021**, *67*, 62. [CrossRef]
53. Zhu, X.; Wang, R.; Shi, W.; Yu, Q.; Li, X.; Chen, X. Automatic Detection and Classification of Dead Nematode-Infested Pine Wood in Stages Based on YOLO v4 and GoogLeNet. *Forests* **2023**, *14*, 601. [CrossRef]
54. Jha, S.; Seo, C.; Yang, E.; Joshi, G.P. Real time object detection and trackingsystem for video surveillance system. *Multimed. Tools Appl.* **2021**, *80*, 3981–3996. [CrossRef]
55. Yin, Y.; Li, H.; Fu, W. Faster-YOLO: An accurate and faster object detection method. *Digit. Signal Process.* **2020**, *102*, 102756. [CrossRef]
56. Meng, W.; Yuan, Y. SGN-YOLO: Detecting Wood Defects with Improved YOLOv5 Based on Semi-Global Network. *Sensors* **2023**, *23*, 8705. [CrossRef]
57. Piazza, M.; Riggio, M. Visual strength-grading and NDT of timber in traditional structures. *J. Build. Apprais.* **2008**, *3*, 267–296. [CrossRef]
58. Cruz, H.; Yeomans, D.; Tsakanika, E.; Macchioni, N.; Jorissen, A.; Touza, M.; Mannucci, M.; Lourenço, P.B. Guidelines for on-site assessment of historic timber structures. *Int. J. Archit. Herit.* **2015**, *9*, 277–289. [CrossRef]
59. Wang, P.; Xiao, J.; Kawaguchi, K.; Wang, L. Automatic Ceiling Damage Detection in Large-Span Structures Based on Computer Vision and Deep Learning. *Sustainability* **2022**, *14*, 3275. [CrossRef]

60. Xu, Y.; Zhang, K.; Wang, L. Metal Surface Defect Detection Using Modified YOLO. *Algorithms* **2021**, *14*, 257. [CrossRef]
61. Rane, N.; Choudhary, S.; Rane, J. YOLO and Faster R-CNN Object Detection in Architecture, Engineering and Construction (AEC): Applications, Challenges, and Future Prospects. Engineering and Construction (AEC): Applications, Challenges, and Future Prospects (October 29, 2023). 2023. Available online: https://ssrn.com/abstract=4624204 (accessed on 30 May 2024). [CrossRef]
62. Weeks, K.D.; Grimmer, A.E. *The Secretary of the Interior's Standards for the Treatment of Historic Properties: With Guidelines for Preserving, Rehabilitating, Restoring & Reconstructing Historic Buildings*; Government Printing Office: Washington, DC, USA, 1995.
63. Bajno, D.; Grzybowska, A.; Bednarz, Ł. Old and Modern Wooden Buildings in the Context of Sustainable Development. *Energies* **2021**, *14*, 5975. [CrossRef]
64. Zhao, L.; Zhi, L.; Zhao, C.; Zheng, W. Fire-YOLO: A Small Target Object Detection Method for Fire Inspection. *Sustainability* **2022**, *14*, 4930. [CrossRef]
65. Zhao, H.W.; Ding, Y.L.; Li, A.Q.; Chen, B.; Wang, K.P. Digital modeling approach of distributional mapping from structural temperature field to temperature-induced strain field for bridges. *J. Civ. Struct. Health Monit.* **2023**, *13*, 251–267. [CrossRef]
66. Zhang, X.; Ding, Y.; Zhao, H.; Yi, L.; Guo, T.; Li, A.; Zou, Y. Mixed Skewness Probability Modeling and Extreme Value Predicting for Physical System Input/Output Based on Full Bayesian Generalized Maximum-Likelihood Estimation. *IEEE Trans. Instrum. Meas.* **2023**, *73*, 1–16. [CrossRef]
67. Zhao, H.; Ding, Y.; Meng, L.; Qin, Z.; Yang, F.; Li, A. Bayesian multiple linear regression and new modeling paradigm for structural deflection robust to data time lag and abnormal signal. *IEEE Sens. J.* **2023**, *23*, 19635–19647. [CrossRef]
68. Katayama, K. Spatial order and typology of Hakka dwellings. In Proceedings of the International Symposium on Innovation and Sustainability of Structures in Civil Engineering, Xiamen, China, 28–30 October 2011; p. 2830.
69. Ma, H.; Li, S. Construction and Application of a System for Assessing the Value of Non-World Heritage Tulou in Pinghe County, Fujian Province, Based on the Analytic-hierarchy Process. *Procedia Environ. Sci.* **2016**, *36*, 114–121. [CrossRef]
70. Ueda, M. A Preliminary Environmental Assessment for the Preservation and Restoration of Fujian Hakka Tulou Complexes. *Sustainability* **2012**, *4*, 2803–2817. [CrossRef]
71. Luo, Y.; Zhong, H.; Ding, N.; Ni, P.; Xu, Y.; Peng, X.; Easa, S.M. Bond–slip mechanism of rammed earth–timber joints in Chinese Hakka Tulou buildings. *J. Struct. Eng.* **2021**, *147*, 04021037. [CrossRef]
72. Chung, C.-H. Dressing Hakka: The Exhibited Luodai Ancient Town. *Glob. Hakka Stud.* **2019**, *11*, 171–190.
73. Lockard, C.A. Chinese migration and settlement in Southeast Asia before 1850: Making fields from the sea. *Hist. Compass* **2013**, *11*, 765–781. [CrossRef]
74. Sullivan, L.F. Traditional Chinese regional architecture: Chinese houses. *J. Hong Kong Branch R. Asiat. Soc.* **1972**, *12*, 130–149.
75. LAI, Y.; Nopudomphan, K. Research on the Architectural Features and Culture from Weiwu of Hakka in Jiangxi. Doctoral Dissertation, Srinakharinwirot University, Bangkok, Thailand, 2023.
76. Aaberg-Jorgensen, J. Clan homes in Fujian. *Arkitekten* **2000**, *28*, 2–9.
77. Luo, Y.; Yang, M.; Ni, P.; Peng, X.; Yuan, X. Degradation of rammed earth under wind-driven rain: The case of Fujian Tulou, China. *Constr. Build. Mater.* **2020**, *261*, 119989. [CrossRef]
78. Liang, R.; Hota, G.; Lei, Y.; Li, Y.; Stanislawski, D.; Jiang, Y. Nondestructive Evaluation of Historic Hakka Rammed Earth Structures. *Sustainability* **2013**, *5*, 298–315. [CrossRef]
79. Ardizzoni, S. The Tulou as a Material Body. In *Hakka Women in Tulou Villages*; Brill: Leiden, The Netherlands, 2022; pp. 82–99. [CrossRef]
80. Xue, L.; Pan, X.; Wang, X.; Zhou, H. Round and Square Buildings and Five-Phoenix Mansions, Ancient Villages in Southwestern Fujian Province. In *Traditional Chinese Villages: Beautiful Nostalgia*; Springer: Singapore, 2021; pp. 73–111. [CrossRef]
81. Wang, L. The logic of peasant resistance to tourism. *Ann. Tour. Res.* **2022**, *97*, 103496. [CrossRef]
82. Hu, L.; Yang, T. Landed and Rooted: A Comparative Study of Traditional Hakka Dwellings (Tulous and Weilong Houses) Based on the Methodology of Space Syntax. *Buildings* **2023**, *13*, 2644. [CrossRef]
83. Shan, D. *Chinese Vernacular Dwellings*; Cambridge University Press: Cambridge, UK, 2011.
84. Lai, W.; Su, X.; Lin, Y. A Preliminary Study on the Current Status of Wooden Structural Materials in Fujian Tulou—Taking Some Tulou in the Ethnic Cultural Village of Hakka Tulou in Yongding as an Example. *Guangdong Build. Mater.* **2010**, *26*, 167–169. [CrossRef]
85. Su, Z.Q. *Exploring the Earthen Buildings*; Central Document Publishing House: Beijing, China, 2006.
86. Ma, H.; Li, S.; Chan, C.S. Analytic Hierarchy Process (AHP)-based assessment of the value of non-World Heritage Tulou: A case study of Pinghe County, Fujian Province. *Tour. Manag. Perspect.* **2018**, *26*, 67–77. [CrossRef]
87. Liu, Y.; Wang, Y.; Dupre, K.; McIlwaine, C. The impacts of world cultural heritage site designation and heritage tourism on community livelihoods: A Chinese case study. *Tour. Manag. Perspect.* **2022**, *43*, 100994. [CrossRef]
88. Shaosen, W.; Jie, H.; Porfyriou, H.; Tournoux, M.N.; Brunori, P. The Fujian Tulou Conservation Strategy: A Sino-Italian joint project. In Proceedings of the ICOMOS—CIAV&ISCEAH 2019 Joint Annual Meeting & International Conference on Vernacular & Earthen Architecture towards Local Development 2019, Pingyao, China, 6–8 September 2019.
89. Wei, S.; Liu, W.; Liang, S. Exploring Hakka Earthen Building Culture and Tourism Development from the Perspective of Protection and Inheritance. *Int. J. Glob. Econ. Manag.* **2024**, *2*, 29–36. [CrossRef]

90. Huang, H.; Huang, H. The Tulou Study: A Historical Review. In *Fujian's Tulou: A Treasure of Chinese Traditional Civilian Residence*; Springer: Singapore, 2020; pp. 1–12. [CrossRef]
91. Sun, Y.; Wang, Z.; Zheng, Y. Environmental Adaptations for Achieving Sustainable Regeneration: A Conceptual Design Analysis on Built Heritage Fujian Tulous. *Sustainability* **2022**, *14*, 11467. [CrossRef]
92. You, X.; Zhang, Y.; Tu, Z.; Xu, L.; Li, L.; Lin, R.; Chen, K.; Chen, S.; Ren, W. Research on the Sustainable Renewal of Architectural Heritage Sites from the Perspective of Extenics—Using the Example of Tulou Renovations in LantianVillage, Longyan City. *Int. J. Environ. Res. Public Health* **2023**, *20*, 4378. [CrossRef] [PubMed]
93. Chen, D.; Su, J.; Ye, J. The Geo-Distribution and Spatial Characteristics of Tulou Dwellings in Chaozhou, Guangdong, China. *Buildings* **2023**, *13*, 2131. [CrossRef]
94. Ueda, M. Environmental assessment and preservation for Fujian Hakka villages. In Proceedings of the International Workshop on Rammed Earth Materials and Sustainable Structures and Hakka Tulou Forum 2011: Structures of Sustainability at International Symposium on Innovation & Sustainability of Structures in Civil Engineering, Xiamen, China, 28–31 October 2011; Xiamen University: Xiamen, China, 2011; pp. 28–31.

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.