DBM-65k: A large-scale multi-scale dataset and benchmark for data-centric bridge damage identification

Junwen Zheng [a], Hao Feng [a], Jinghuan Zhang [a], Jian Zhang [a, b, c, *]

[a] School of Civil Engineering, Southeast University, 210096, Nanjing, China

[b] Advanced Ocean Institute of Southeast University, 226000, Nantong, China

[c] Suzhou Research Institute of Southeast University, 215000, Suzhou, China

* Correspondence and requests for materials should be addressed to jian@seu.edu.cn (Jian Zhang)

Abstract: UAV bridge damage inspection has long been limited by insufficient data scale, a single damage type, and lack of unified training paradigm for the model, making it difficult to achieve ideal results in real scenarios. This paper constructed the DBM-65k large-scale benchmark dataset and proposed the DBM (Detection Bridge Model) unified architecture model series to establish a comprehensive evaluation benchmark for bridge detection. Based on more than 65,000 images, DBM-65k constructed multi-scale and multi-category scenes, establishing a hierarchical perception system of "macro subject appearance and general micro-components". This paper uses YOLOv12 as the baseline, designs the dynamic alignment fusion module (DAF) and the dynamic detection target head (P2Head), and introduces adaptive scale-aware loss to amplify the characteristics of small-scale damage by dynamically adjusting the gradient return. For damage segmentation, a detection prior-driven cross-task transfer learning strategy is adopted to achieve collaborative training of detection and segmentation tasks. Experimental results show that the DBM series models show good performance in both macro-disease and micro-component tasks, especially achieving a significant gain of up to 5.37% in small object detection accuracy (mAPs). At the same time, DBM-seg provides high-precision pixel-level segmentation output. The training weights have been open sourced. This work will provide a solid data and algorithm basis for unified bridge structure detection.

Keywords: Bridge damage detection; multi-scale design; multi-type bridges; computer vision; YOLOv12

Highlights：

1. A dynamic collection strategy was used to construct the bridge damage benchmark dataset DBM-65k, which covers the multi-scale category characteristics of bridge damage;
2. Developed a DBM model integrating the dynamic alignment fusion module and the high-resolution P2 detection head to improve the damage detection rate;
3. Designed an adaptive scale-aware loss function (Adaptive Scale-aware Loss) to amplify the gradient response of small-scale damage;
4. Proposed detection prior-driven cross-task transfer learning to achieve collaborative optimization of detection and segmentation.

1. Introduction

As a large number of bridges gradually enter the aging stage, bridge damage problems caused by material deterioration, environmental erosion and long-term loading have become increasingly prominent, seriously threatening the durability and operational safety of bridge structures.[1, 2]. Therefore, realizing high-efficiency and high-precision intelligent bridge inspection has become an important research direction in the field of structural health inspection.[3]. Traditional manual inspections have problems such as low efficiency, high risks, and limited coverage, making it difficult to meet the inspection needs of bridges.[4]. In contrast, unmanned aerial vehicles (UAVs) have gradually become an

important technical means for intelligent bridge inspection due to their advantages such as high mobility, flexible deployment and the ability to quickly acquire images.[5, 6]. At the same time, the development of deep learning has further promoted the application of drone visual inspection technology in bridge engineering.[7].

However, there are significant differences between bridge types, disease forms, component scales, and shooting angles, which results in the generalization ability of existing methods in complex scenes being still limited.[8]. Drone bridge inspection requires not only the superficial inspection of the structure, but also the refined identification of components. Therefore, the core issue of intelligent bridge detection is no longer limited to model structure optimization, but how to build a large-scale data system and unified detection framework that adapts to real inspection scenarios.[9].

1.1 Bridge drone inspection challenges

The real bridge inspection environment has typical complex engineering attributes, and its core difficulties mainly come from the diversity of bridge types, large target scale spans, and complex background interference.[10].

Bridge inspection tasks have obvious multi-type and multi-level characteristics. Different bridges have significant differences in structural form, material system and service environment: concrete bridges pay more attention to superficial diseases such as cracks, holes, spalling and exposed bars, while steel structure bridges mainly focus on rust, coating failure and corrosion expansion.[11]. During the actual engineering inspection process, it is necessary to further conduct refined status inspections of key components such as cables, bolts, and connecting nodes.[12]. Z Yao et al.[13]A multi-scale defect fusion segmentation network is proposed to achieve collaborative detection and high-precision edge segmentation of concrete cracks and depressions, emphasizing the necessity of multi-level and multi-object detection in real inspections. E Figueiredo et al.[11]The impact of climate change on bridge structure damage detection was discussed, and it was pointed out that bridges with different material systems have essential differences in disease manifestations and detection needs. This means that bridge detection is not a single category identification task, but a comprehensive task covering both macroscopic structural damage detection and microscopic key component diagnosis. Different diseases have significant differences in texture features, geometric shapes, and spatial distribution, making it difficult for the model to establish a unified and stable feature expression capability.[14].

UAV bridge inspection scenarios commonly suffer from significant scale changes and complex environmental interference problems. Therefore, in the same inspection mission, both large-scale targets and long-distance microscopic diseases may appear.[7]. For example, targets such as fine cracks, bolt corrosion, and cable damage are small in size under long-distance shooting conditions, which can easily cause feature information to be lost during the deep network downsampling process; while large-scale spalling and steel structure corrosion areas shot at close range are usually accompanied by complex boundaries, irregular shapes, and large scale spans.[15]. M. Khan et al.[16]In the review, the key role of multi-scale feature fusion in bridge damage detection was explained, and the inherent limitations of the traditional fixed-scale feature extraction structure in dealing with targets with extreme scale differences were pointed out. T. Liu et al.[17]A bridge crack segmentation method based on densely connected U-Net is proposed. By introducing multiple types of modules, the model's ability to capture long-range dependence information of small cracks in complex backgrounds is improved. This extreme scale difference makes it difficult for traditional fixed feature fusion structures to take into account both global semantic expression and tiny target detail perception capabilities.[18].

In addition to the challenges at the visual level, bridge inspection also has obvious data heterogeneity issues. Different bridges differ greatly in terms of structural scale, disease distribution, and shooting conditions. The same disease may also present completely different visual characteristics in different environments. For example, concrete cracks will be affected by shooting distance, lighting angle, and surface contamination, while steel structure corrosion may show different colors and textures as the service environment changes.[19-21]. M. Maguire et al.[22]The released dataset contains more than 56,000 binary classification images of concrete cracks and non-cracks and contains a variety of interference conditions, providing a basis for the benchmark evaluation of crack classification algorithms. R.-s. Ji et al.[23]A lightweight bridge surface defect detection network based on YOLOv10 is proposed. By introducing lightweight structures, multi-scale feature enhancement and model compression are achieved. Y. Gao et al.[24]From the perspective of few-sample learning, a small-sample learning method is proposed. Through cross-domain transfer learning and embedding normalization technology, high-precision bridge damage detection can be completed with only a small number of labeled samples. However, in actual application environments, this highly complex data distribution further increases the difficulty of model generalization.[25].

Therefore, UAV bridge inspection is a comprehensive visual perception problem that integrates multi-type bridges, multi-level components, multi-scale targets and complex background interference. How to build a large-scale data system that can cover the real inspection environment and establish a unified detection framework that takes into account both macrostructure recognition and microdetail perception capabilities has become an important issue that needs to be solved urgently in current bridge intelligent inspection research.

1.2 Limitations of existing research

Since the advent of AlexNet in 2012[26], research on damage identification based on computer vision is emerging one after another, but research in this field still faces severe structural limitations. Y.J. Cha et al.[27]The early application of CNN to the detection of concrete damage types demonstrated the great potential of deep learning networks in the field of civil engineering, but its research also revealed the network's high dependence on the scale and diversity of training data. The vast majority of existing research still stays in the model-driven paradigm, trying to improve the indicators on a specific experimental set by adjusting the network structure, introducing different attention mechanisms, or optimizing hyperparameters.[28-32]. However, this strategy often still shows a serious failure tendency when faced with bridge detection tasks in real, open scenarios. The fundamental reason is that existing research does not pay enough attention to the breadth, depth and structure of the underlying dataset, resulting in a lack of prior knowledge of the complex damage characteristics of the real world in the model.[33-35].

The current public datasets show obvious fragmentation characteristics. Although mainstream datasets promote the development of algorithms on specific tasks, their data size is usually limited to hundreds or thousands of images, and the collection scenarios are mostly ideal laboratory environments or specific single bridges.[36-39]. Such datasets often only label a specific disease (such as a single concrete crack), ignoring the coexistence of multiple types of damage to concrete, steel, cables, bolts, etc. in actual projects. This disconnect between training data and real business scenarios makes it difficult to guarantee the robustness of the model when dealing with heterogeneous bridges or cross-material tasks. V. Giglioni et al.[40]The application of domain adaptation technology in bridge damage detection was studied, and it was found that when the bridge structural form or environmental conditions change, the performance of the traditional transfer learning method degrades sharply, which further confirms the key

role of training data diversity.

Although multi-scale feature fusion networks (such as FPN) have been widely used in object detection tasks[41], but in the scenario of long-distance drone inspection, targets with small-scale damage are still prone to feature loss. Since small diseases usually occupy only a few pixel areas, their texture and edge information are easily weakened during the continuous downsampling process, resulting in insufficient effective disease features in deep features. In addition, the current field of intelligent bridge detection still lacks unified and public large-scale pre-training weights and standardized evaluation benchmarks, which to a certain extent limits the development of a universal bridge damage detection framework.

Faced with these limitations, the field of bridge disease identification urgently needs to shift to a data-centered paradigm, and the research focus should shift from simple model construction to the construction of ultra-large-scale, spatially hierarchical benchmark datasets covering multiple types of damage.[42]. By building a data base with large-scale real-world distribution, supplemented by a unified training framework that can adaptively adjust scale perception, we can truly solve the difference between the current laboratory model and engineering reality, and realize the leap from single task recognition to full-scenario bridge detection.[43].

1.3 The core contribution of this article

In view of the problems in existing research such as limited data scale, fragmented task systems, insufficient multi-scale target sensing capabilities, and weak openness, this study is data-centered and builds a large-scale data and unified detection framework around the complex engineering scenarios of real drone bridge inspections. The main contributions of this article are as follows:

(1) Based on the dynamic acquisition strategy, a bridge damage benchmark dataset DBM-65k was constructed, covering multi-scale category characteristics of bridge damage;

(2) A DBM model integrating the dynamic alignment fusion module and the high-resolution P2 detection head was developed to improve the damage detection rate;

(3) An adaptive scale-aware loss function was designed to amplify the gradient response of small-scale damage;

(4) Proposed detection prior-driven cross-task transfer learning to achieve collaborative optimization of detection and segmentation.

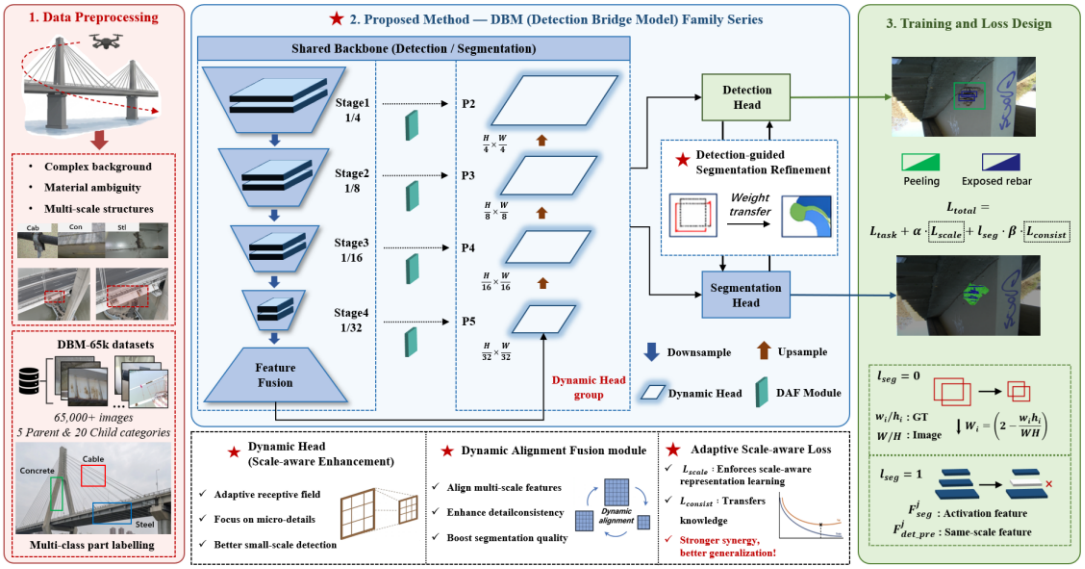

Figure 1. DBM-65k and DBM research roadmap

2. DBM-65k dataset

Although existing bridge damage datasets have promoted algorithm development, there are still significant limitations (see Table 1). SDNET2018 is limited to the second classification of short-range cracks; CODEBRIM is too small, with only 1,590 images; although dacl10k and GYU-DET introduce multi-defect annotation, the data size is less than 10,000 images and is limited to a single material of concrete. More importantly, existing data generally lack real UAV multi-scale scenes, making it difficult to coordinate the coordinated detection of large-scale structural appearance and microscopic defects.

In order to break through the above bottlenecks of limited scale, single scenes and materials, and lack of multi-scale features, this paper constructed the DBM-65k large-scale benchmark dataset. This dataset is designed with the concept of "macro-micro hierarchical perception" to truly restore the complex drone inspection engineering environment. It has achieved comprehensive breakthroughs in data volume, disease diversity, extreme scale span and comprehensiveness of assessment tasks, and established a new industry benchmark.

Table 1. Comparison table of mainstream public bridge datasets

Dataset	Number of images	damage/target type	bridge type	Annotation type	Main features and limitations
SDNET2018[44]	56,000	Cracks (two categories: cracks/no cracks)	Concrete (bridge deck, wall, pavement)	Image classification	Large scale, but single task, single material, no multi-category detection/segmentation
CODEBRIM[45]	1,590	Category 6 (cracks, peeling, exposed tendons, efflorescence, rust, non-defects, etc.)	concrete bridge	Object detection	The diversity of real bridges is good, but the scale is small and the processing of overlapping defects is complicated.
dacl10k[46]	9,920	**19 categories (13 types of damage + 6 types of components)**	concrete bridge	Semantic segmentation	The diversity of real scenes is strong, but the subcategories are small and there are no steel structures/micro components
GYU-DET[47]	11,123	Category 6 (cracks, spalling, water seepage, honeycomb,	concrete bridge	Object detection	Multi-category defect dataset, but limited to concrete girder

		exposed tendons, holes)			bridges and single materials
COCO-Bridge[48]	774	Bridge structural components (non-damage)	universal bridge	Object detection	Focus on component detection rather than surface damage, damage correlation is weak
Corrosion Condition Rating Database[49]	514	Category 3 (corrosion severity)	steel bridge	Semantic segmentation	The data size is small and the scenes and materials are single.

2.1 Data collection strategy

The data collection of DBM-65k relies on real drone inspection missions and adopts a multi-height, multi-view dynamic flight strategy to highly restore the visual distribution characteristics of complex engineering scenes (Figure 1). In terms of bridge and damage types, the dataset comprehensively covers concrete bridges (such as box girders, continuous beams) and steel structure bridges (such as steel trusses, connecting nodes). The annotation range not only includes conventional surface defects such as cracks, spalling, and rust, but also extends to the microscopic level to identify cable damage and multi-state abnormality of bolts, which is highly consistent with real comprehensive inspection needs.

In addition, in order to enhance the diversity of target scales, the acquisition process deeply integrates long-distance macrostructure cruise and close-up local component close-ups, so that the dataset includes both large-scale structural apparent diseases and long-distance micro-targets. In the end, DBM-65k built a large-scale real data base that integrates multiple bridge types, multiple diseases, extreme multi-scales and complex background interference, laying a solid data foundation for the subsequent development of a unified detection architecture.

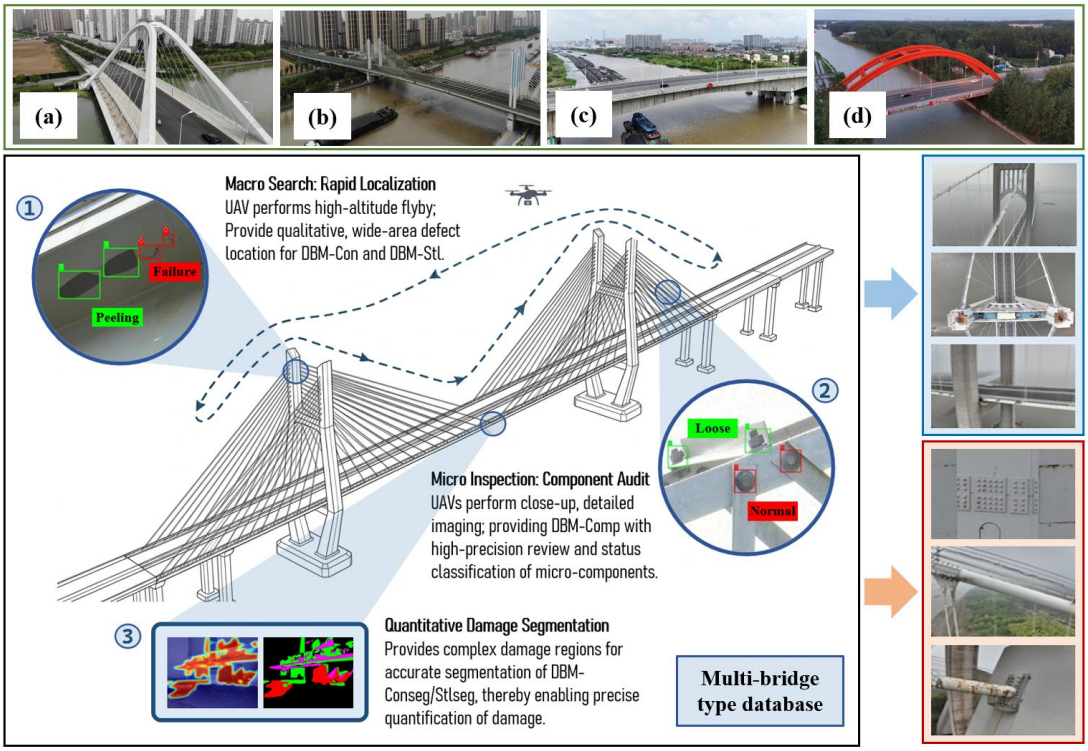

Figure 2. Diverse data acquisition scenarios and multi-source collection.

2.2 Macro-micro layered system

Different from the traditional single-task bridge disease dataset, DBM-65k uses a macro-micro hierarchical organization method to uniformly manage different tasks to be more consistent with the engineering logic in real bridge inspections. The names and specific classifications of the sub-datasets are shown in Table 2.

The macro layer is mainly oriented towards the detection and segmentation of bridge main structure diseases, focusing on large-scale structural apparent diseases such as concrete cracks, spalling, exposed bars and steel structure corrosion. Such targets usually have obvious regional characteristics and place more emphasis on overall structural condition assessment capabilities.

The microscopic layer is mainly targeted at key connection components and long-distance fine-grained defect detection tasks, including bolt status detection, cable damage identification, etc. Such targets are usually smaller in size, have weaker textures, and are more susceptible to complex backgrounds and scale changes. Therefore, higher requirements are placed on the model's detail perception capabilities.

This macro-micro layered system can not only effectively describe the task hierarchical relationships in real bridge inspections, but also provide a clear data organization structure for subsequent unified model design. Compared with traditional single-category datasets, DBM-65k highlights the need for cross-level collaborative detection in bridge inspections, thereby more truly reflecting visual perception problems in complex engineering scenes.

Table 2. DBM-65k macro and micro task system and category definitions

Model name	Assessment level	Task type	Applicable objects	Breakdown of Injury and Condition Categories	core positioning in the system
DBM-	Macro-	Object	concrete	Cracks, peeling, rust,	Rapid detection

Con	scale appearance	detection	structure	holes	and location of concrete surface damage in images
DBM-Conseg	Macro-scale appearance	Image segmentation	concrete structure	Cracks, efflorescence, peeling, exposed tendons, holes	Pixel-level precise extraction of complex apparent disease geometry
DBM-Stl	Macro-scale appearance	Object detection	steel structure	Early coating failure, coating peeling, and surface corrosion	Rapid detection and positioning of steel structure bridge damage in images
DBM-Stlseg	Macro-scale appearance	Image segmentation	steel structure	Rust (rated by degree of deterioration: fair, poor, severe)	Pixel-level precise extraction of corrosion diseases of steel structure bridges
DBM-Comp	microscopic components	Object detection	member	Cable (broken); Bolt (normal, missing, rusty, loose)	Detailed investigation of micro-node status

2.2.1 Macrostructure apparent data

At the macro level, DBM-65k is mainly oriented to the task of detecting and segmenting bridge main structure diseases, covering the apparent status of concrete and steel structure bridges, as shown in Figure 3. The concrete body sub-dataset includes DBM-Con and DBM-Conseg. DBM-Con focuses on object detection tasks and covers four typical diseases: cracks, spalling, rust and holes. It aims to quickly identify bridge surface damage and locate their spatial distribution in images. DBM-Conseg is further oriented towards image segmentation tasks, and can annotate complex superficial diseases such as cracks, efflorescence, spalling, exposed tendons and holes at the pixel level to achieve precise extraction of geometric shapes of the diseases. This segmentation-level annotation can provide the model with richer local texture and boundary information, thereby improving the identification accuracy of complex cracks and surface peeling areas.

The steel structure main subdataset includes DBM-Stl and DBM-Stlseg. Among them, DBM-Stl is targeted at object detection tasks, covering three types of diseases: early coating failure, coating peeling and surface corrosion, and is used to locate damaged areas on the surface of steel structures; DBM-Stlseg provides pixel-level segmentation and annotation of rust diseases, and divides them into three levels: general, poor and severe according to the degree of deterioration, providing quantitative analysis of the corrosion status of steel structure bridges.

Through macro-level organization, DBM-65k can truly reflect the large-scale visual characteristics of the main bridge structure during drone inspections, while taking into account the unified assessment needs of different material systems and damage types.

2.2.2 Microscopic general component data

At the micro level, DBM-65k focuses on the precise detection of key bridge components and long-

distance defects. The micro-component sub-dataset DBM-Comp covers multi-state detection of cable damage and bolts. The bolt status is divided into four categories: normal, rusted, loose and missing, and the cables are marked according to the damage status, as shown in Figure 3. Since these tiny targets often only occupy a small number of pixel areas in long-distance drone inspection images, DBM-Comp's annotation strategy combines high-resolution local magnification inspection and fine bounding box annotation to ensure that subtle targets are fully characterized during training.

Micro-level data not only involves material differences, but also takes into account component functional differences: bolts and cables are core components of bridge connection nodes, and their status is directly related to the safety of the overall structure. Therefore, high-precision, fine-grained object detection and status classification put forward higher requirements for the model. DBM-Comp provides a standardized data basis for cross-component type micro-target recognition through unified annotation and multi-state classification. It also supports multi-task models to jointly learn and optimize macroscopic appearance information and microscopic component states.

Microscopic universal component inspection can make up for the fine-grained defect information that cannot be covered in macro-level evaluation, allowing the model to maintain high recognition accuracy even at long distances, small targets, and complex backgrounds. Through the combination of macro-micro layering, DBM-65k achieves full scene coverage from the apparent state of the overall bridge to the fine defects of key components, providing comprehensive data support for the end-to-end application of drone intelligent bridge inspection.

Figure 3. Representative samples across five task categories.

2.3 Statistical analysis of dataset characteristics

In order to fully reflect the scale, category distribution and data annotation quality of DBM-65k, this section provides a detailed description from four aspects: statistical characteristics, category distribution, target size characteristics and annotation quality control process. Through strict statistical analysis and multi-level quality control, DBM-65k not only ensures the diversity of large-scale data, but also provides a reliable training and evaluation basis for macro and micro tasks.

DBM-65k contains a total of more than 65,000 images and nearly 300,000 annotated instances, covering multiple types and multi-level scenarios such as concrete bridges, steel structure bridges, and key connecting components. Detailed statistical information is shown in Table 3. Among them, the macrostructure appearance evaluation sub-dataset (DBM-Con, DBM-Conseg, DBM-Stl, DBM-Stlseg) has a total of about 40,000 images, including six types of macroscopic diseases: cracks, spalling, holes, exposed tendons, coating failure and corrosion. The microscopic common component fine inspection

subdataset (DBM-Comp) contains approximately 25,000 images, covering two types of fine-grained targets: bolt status (normal, rusty, loose, missing) and cable damage.

The distribution of instances of each category is relatively balanced at the macro level and micro level to ensure that the model can learn sufficient multi-category features during the training process. The macro layer cracks and spalling instances have the largest number, accounting for approximately 55% of the macro sub-dataset, steel structure corrosion and coating failure account for approximately 30%, and the remaining categories account for approximately 15%. Microscopic layer bolt defects and cable damage account for approximately 70% and 30% of the microscopic sub-dataset respectively, ensuring that fine-grained targets can be fully learned even under long-distance shooting.

Table 3 Detailed statistics of DBM-65k subset

Parent category	Images	Instances (parent)	Parent category percentage (%)	Child category	Instances (child)	Child category percentage (%)
DBM-Con	40,576	168,972	56.74%	Corrosion	59,645	20.03%
				Peeling	39,001	13.10%
				Crack	36,123	12.13%
				Hole	34,203	11.48%
DBM-Conseg	7,202	30,432	10.22%	Hole	13,501	4.53%
				Efflorescence	8,587	2.88%
				Peeling	3,474	1.17%
				Crack	2,837	0.95%
				Exposed rebar	2,033	0.68%
DBM-Stl	5,185	17,890	6.01%	Corrosion	14,747	4.95%
				Coating failure	2,129	0.71%
				Coating peeling	1,014	0.34%
DBM-Stlseg	2,862	15,177	5.10%	Fair	10,623	3.57%
				Poor	3,933	1.32%
				Severe	621	0.21%
DBM-Comp	10,021	65,341	21.94%	Normal (bolt)	36,685	12.32%
				Rust (bolt)	6,414	2.15%
				Loose (bolt)	535	0.18%
				Missing (bolt)	248	0.08%
				Broken (cable)	21,459	7.21%
Total	65,846	297,812	100%	—	297,812	100.00%

2.3.1 Multi-scale object detection frame

The object detection sub-dataset of DBM-65k includes two categories: macrostructure apparent diseases (DBM-Con, DBM-Stl) and microscopic key components (DBM-Comp). In the macro level, the relative areas of concrete and steel structure targets are mostly concentrated in the small to medium range. The relative areas of cracks, spalling and holes in DBM-Con are mainly less than 5%, while the relative areas of corrosion and coating failure of DBM-Stl steel structures are more widely distributed, extending from less than 5% to about 50%, reflecting the high heterogeneity of the areas covered by steel structure

diseases in drone inspection images. The target sizes of the microstructure subdataset (DBM-Comp) are generally small, and the relative area is mostly less than 1%, emphasizing the significant sparseness of small targets under long-distance shooting conditions.

The target size heat map shows that the macroscopic targets are relatively evenly distributed in aspect ratio. The target widths and heights of DBM-Con and DBM-Stl are relatively concentrated in the range of 0.2 to 0.6, indicating that most diseased areas are long or medium square, while micro-component targets tend to be in low aspect ratio areas. The vertical proportion of bolts and cable targets is slightly higher than the horizontal ratio, reflecting their long strip characteristics. These statistical characteristics reveal the differences in scale and shape of targets in UAV bridge inspection, providing a design basis for multi-scale feature extraction and dynamic scale alignment.

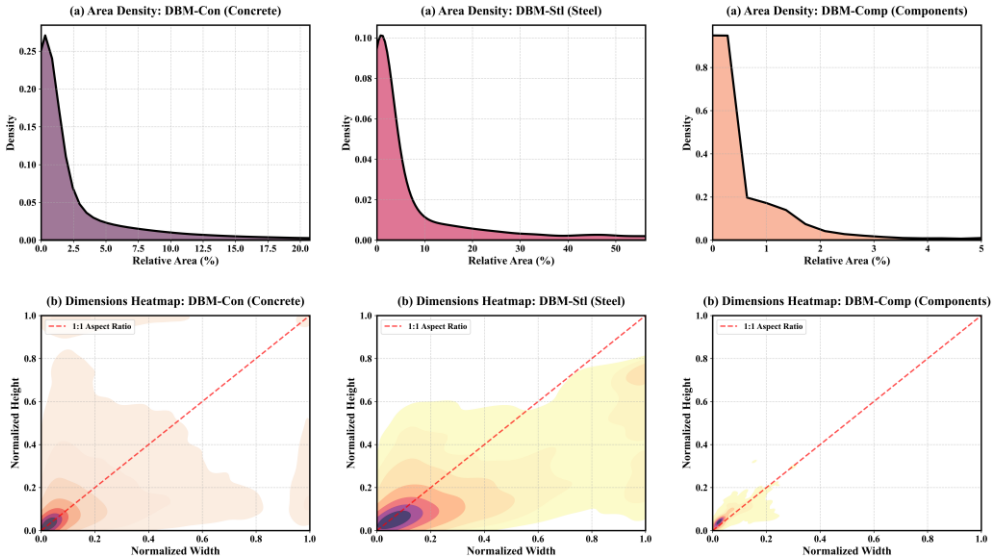

Figure 4. Target relative area probability density distribution map and bounding box normalized dimension representation map.

2.3.2 Polygon segmentation annotation

Segmentation sub-datasets (DBM-Conseg, DBM-Stlseg) provide pixel-level disease annotations, covering categories such as concrete cracks, spalling, exposed bars, and steel structure corrosion. Statistics on the number of vertices show that the diversity of macroscopic crack and spalling annotations is high, and the number of vertices is widely distributed. The average number of vertices of spalling annotations is about 300–400, and the number of crack vertices can exceed 600, reflecting complex boundary and irregular morphological characteristics. The number of vertices of steel structure corrosion annotation (DBM-Stlseg) is concentrated between 150 and 300, and the number of vertices in severe damaged areas is larger, reflecting the increase in disease area and boundary complexity.

Effective pixel ratio statistics further reveal the proportion of different categories in the image. The proportion of macro cracks and spalling annotated pixels is low, but it provides rich edge and texture information in the overall training; the effective proportion span of the steel structure corrosion category is large, extending from about 30% to nearly 80%, which fully reflects the difference in disease coverage. These statistical analyzes provide model designers with a quantitative basis for macro and micro segmentation tasks, helping to optimize loss functions, feature fusion and multi-scale processing strategies.

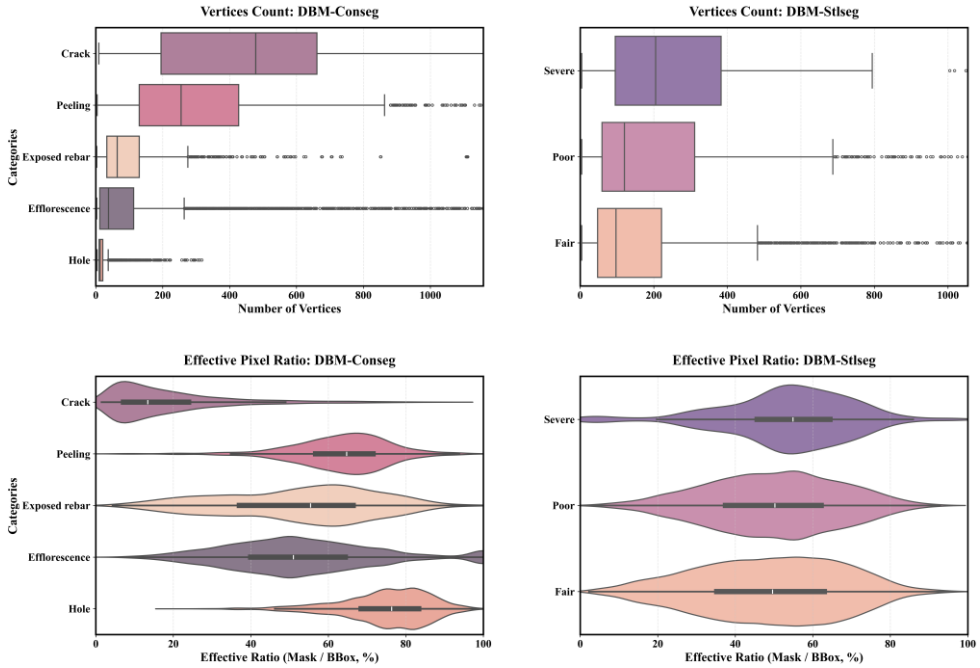

Figure 5. Annotation complexity statistics and effective pixel proportion distribution chart.

3. DBM bridge detection model

In order to effectively deal with the challenges of large macro-micro span, strong background interference, and complex damage topology presented in the DBM-65k dataset, this paper proposes a unified visual perception architecture called DBM. This model series achieves the identification of a unified framework for macro-structural diseases and micro-component defects by introducing multi-scale dynamic feature alignment, fine-grained micro-target perception, and cross-task knowledge transfer mechanisms.

3.1 Multi-tasking unified infrastructure

Complex UAV inspection tasks need to take into account both reasoning efficiency and detection accuracy. The DBM model series uses YOLOv12 as a powerful and versatile baseline architecture.[50]. With its advanced area attention mechanism (Area Attention) and efficient feature aggregation network, YOLOv12 has demonstrated excellent feature extraction capabilities in general computer vision tasks. In the DBM architecture, the backbone network of YOLOv12 mainly serves as the base for feature extraction. Through a multi-stage down-sampling process (Stage 1 to Stage 4), it gradually extracts shallow high-resolution features containing rich local textures to low-resolution features containing deep global semantics.

However, the native YOLOv12 architecture is designed primarily for general-purpose scaling goals. When faced with long-distance bridge inspections by UAVs, the original neck network (Neck) and detection head (Head) lack an accurate spatial alignment mechanism when processing feature cross-layer fusion, and the default maximum downsampling step size can easily lead to insufficient feature extraction of targets such as tiny bolts and fine cracks in the deep network.[51-53]. To this end, based on the powerful feature base of YOLOv12, this paper conducts in-depth architecture customization and optimization for the multi-scale damage characteristics of bridges.

3.2 Multi-scale model architecture optimization

Real UAV bridge inspection faces multi-scale span challenges. The targets include not only large-area superficial diseases covering the surface of the structure, but also small defects in remote components occupying a small number of pixels. When conventional detection models deal with such

drastic scale changes, their default large-step downsampling mechanism can easily lead to insufficient feature extraction and information collapse of underlying features of tiny cracks or tiny bolts in deep networks. At the same time, in the traditional feature pyramid network (FPN), when deep low-resolution semantic features and shallow high-resolution spatial features are fused across layers, there is often a lack of precise spatial alignment mechanism, which can easily lead to feature misalignment and background noise aliasing.

In response to the above core pain points of difficulty in cross-scale feature fusion and loss of small target features, the DBM architecture has been deeply customized and optimized in the FPN structure, and innovatively introduced the dynamic alignment fusion module (DAF) and the dynamic small object detection head (Dynamic P2 Head), as shown in Figure 6. This combination mechanism reconstructs the network's collaborative sensing capabilities for bridge macro- and micro-heterogeneous diseases through spatial adaptive alignment of multi-scale features and dynamic enhancement of underlying high-frequency micro features.

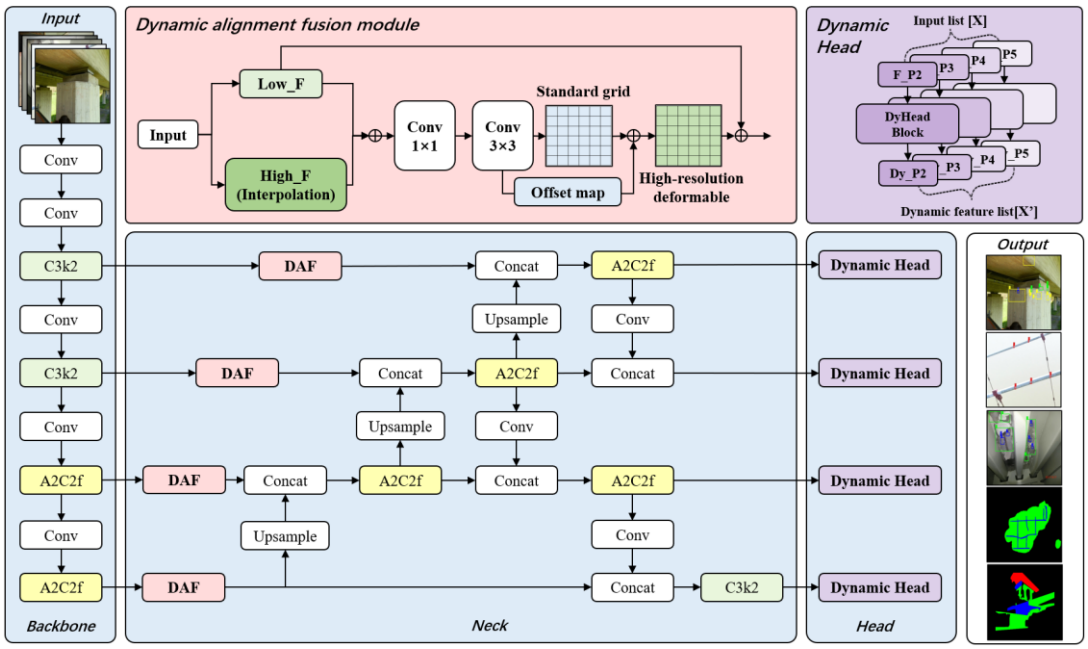

Figure 6. The unified architecture of the DBM series.

3.2.1 Dynamic Alignment Fusion Module (DAF)

In the multi-scale feature fusion process, the shallow feature map retains rich spatial geometry and high-frequency edge details, while the deep feature map undergoes multiple pooling and convolution superpositions, and its receptive field is expanded to extract a highly abstract global semantic representation. Since there is information loss when the two experience different downsampling steps, semantic misalignment occurs when deep features are mapped to shallow layers. Especially in bridge inspection scenarios, disease targets span large scales and have complex geometric topology. Traditional bilinear interpolation upsampling is followed by splicing or element-by-element addition, which essentially performs feature coupling under fixed regular grid constraints. This fusion mechanism will not only cause serious feature aliasing and geometric distortion, but also introduce high-frequency background noise in deep features into shallow high-resolution feature space.

In order to enhance the semantic balance of far and near features and actively suppress complex background noise, this paper designs a dynamic alignment fusion module (DAF), as shown in Figure 7.

The core theoretical basis of DAF is to use deformable convolution to construct an adaptive spatial geometric transformation field and complete pixel-level semantic center alignment before cross-layer feature fusion.[54].Specifically, given deep low-resolution features rich in global semantic information F_{high} and shallow high-resolution features rich in fine-grained details F_{low}, the module first F_{high} Do preliminary upsampling to match F_{low} The spatial resolution, then, the network implicitly predicts the offset matrix (Offset map) describing the spatial deformation trend by jointly learning these two heterogeneous features. ΔP：

$$\Delta P = Conv_{3\times3}(Conv_{1\times1}(F_{low} \oplus Upsample(F_{high}))) \quad (1)$$

in, $\oplus$ Represents element-wise addition, aiming to build the initial semantic interaction context of deep and shallow features; $Conv_{1\times1}$ Used for cross-channel feature dimensionality reduction and information compression, $Conv_{3\times3}$ As an offset field generator, it outputs a two-dimensional continuous spatial offset vector with a channel number of 2K.

Getting the spatial offset ΔP Finally, the DAF module performs spatial resampling and adaptive deformation operations at high resolution on the deep feature grid, so that the deep semantic features actively move closer to the real geometric edge and local texture center of the shallow disease. Its mathematical expression is:

$$F_{aligned}(p) = \sum_{k=1}^{K} w_k \cdot Upsample(F_{high})(p + p_k + \Delta p_k) \quad (2)$$

In the formula, p is the reference pixel position on the feature map, K is the total number of effective sampling points of the convolution kernel, w_k The feature aggregation weight of the corresponding position, p_k is the local sampling offset under the standard regular grid, Δp_k It is the continuous dynamic space offset learned by the model adaptively. because Δp_k Usually containing sub-pixel level decimal places, the system accurately calculates the characteristic response value in its continuous space through a bilinear interpolation algorithm. Through the DAF module, the network can effectively solve the geometric distortion problem in the feature fusion process, significantly improve the response intensity of the damage boundary, and lay the foundation for subsequent high-quality detection and fine segmentation.

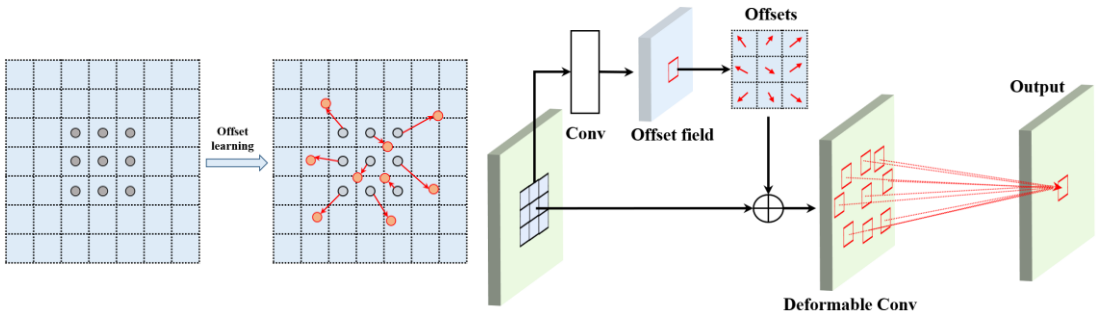

Figure 7. Offset calculation and high-resolution deformable.

3.2.2 Dynamic small object detection head (Dynamic P2 Head)

The general object detection model usually relies on the feature pyramid levels of P3, P4, and P5 scales (corresponding to 1/8, 1/16, and 1/32 downsampling respectively) for target prediction. This design performs well when processing conventional natural scene datasets, but is limited by the safe flight distance of the drone, and the pixel area of microscopic targets such as bolts and broken cables is often small. This type of tiny target undergoes continuous downsampling pooling operations such as 8

times (P3 layer), and its response area on the deep feature map will shrink to the sub-pixel level (Sub-pixel). This will not only lose the edge texture details of the target, but also cause its weak activation signal to be masked by complex background noise, resulting in a large number of missed detections.

To this end, the DBM architecture specifically customizes a high-resolution P2 detection layer (downsampling step size is 4, size is $\frac{H}{4} \times \frac{W}{4}$) to preserve the microscopic details of the underlying image to the maximum extent. However, there is an inherent theoretical flaw in directly introducing shallow high-resolution feature maps: although the shallow network has high spatial positioning accuracy, its receptive field is limited, lacks deep global semantic information, and is filled with a large amount of high-frequency background noise. If the P2 layer is directly used for prediction, it will easily lead to a large number of false positives and false alarms.

In order to effectively aggregate these high-frequency microscopic features and actively suppress the accompanying background noise, this article introduces a dynamic head mechanism (DyHead Block) in front of each layer. This module uses scale-aware and spatial-aware attention mechanisms in series to analyze input features. $F \in \mathbb{R}^{C \times H \times W}$ Perform refined recalibration.

First, there are significant differences in the activation responses of small targets between different channels, and the scale-aware module aims to suppress noise channels through adaptive weighting of feature channels. This module first extracts the global context distribution of the channel dimension through global average pooling (GAP), generates scale calibration weights including environmental priors, and then performs preliminary channel reconstruction of the features:

$$F_{scale} = F \otimes \sigma(W_2 \cdot \delta(W_1 \cdot GAP(F))) \tag{3}$$

In the formula, $\otimes$ Represents element-wise multiplication along the channel dimension; $GAP(\cdot)$ It is the global average pooling operation; $W_1 \in \mathbb{R}^{\frac{C}{r} \times C}$ and $W_2 \in \mathbb{R}^{C \times \frac{C}{r}}$ are respectively the dimensionality reduction and dimensionality enhancement weight matrices of the multilayer perceptron (r is the channel scaling); δ is the ReLU activation function, σ is the Sigmoid normalization function. Through formula (3), the network can adaptively attenuate the channel weights that are sensitive to the concrete background, while simultaneously stimulating feature channels that are highly sensitive to rust color or metal edges.

In order to further enhance the spatial sensitivity of the model to local small defects, the spatial perception module F_{scale} The spatial topology information was refined. For example, determining whether a bolt is "loose" is highly dependent on the geometric gap of a few pixels between the nut and the washer. In order to accurately capture this weak spatial gradient change without introducing excessive computational burden, this module uses depthwise separable convolution (DW Conv) for spatial modeling:

$$F_{out} = F_{scale} \otimes \sigma(Conv_{1\times1}(DWConv_{3\times3}(F_{scale}))) \tag{4}$$

In the formula, $DWConv_{3\times3}$ Represents a 3×3 depthwise separable convolution designed to efficiently capture the local geometric response of the disease; $Conv_{1\times1}$ It is used for cross-channel information exchange. Through the forward pass of these two formulas, the output features F_{out} Dual alignment in multi-scale and local spatial dimensions is achieved. The introduction of Dynamic P2 Head balances the contradiction between high-resolution feature extraction and high computing power costs. It provides a network layer solution to the problem of missed detection of small defects caused by long-distance drone inspections with almost no additional computing overhead.

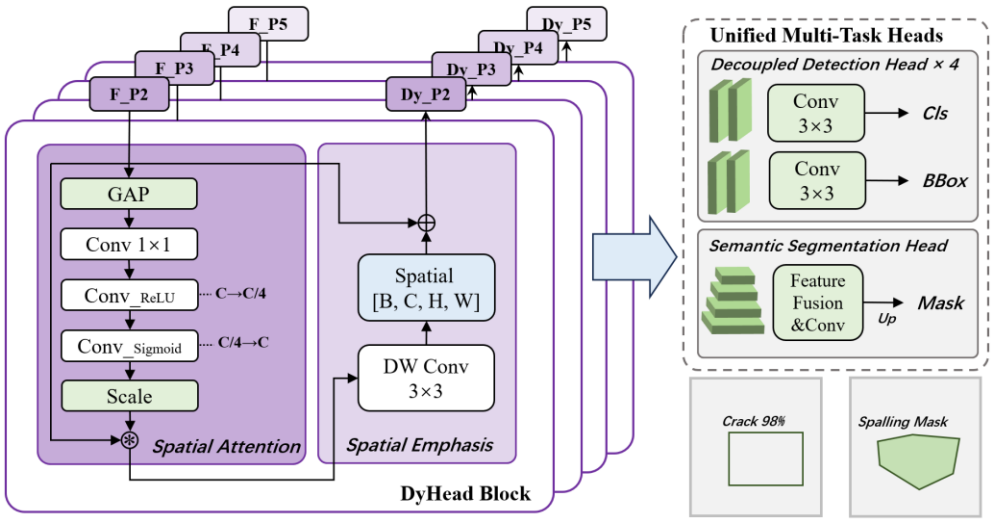

Figure 8. Multi-dimensional dynamic combination detection head.

3.3 Task-oriented training strategy

Real bridge inspection requires both rapid location of diseases (object detection) and morphological quantification of complex damage (semantic segmentation). In order to efficiently collaborate these two tasks within a unified theoretical framework, this paper designs a task-oriented multi-scale consistency loss function and proposes a detection prior-driven cross-task transfer learning strategy.

3.3.1 Perceptual loss of macro and micro detection scales

In order to enable the network to adaptively take into account macroscopic diseases and microscopic small defects during the optimization process, this paper constructs a joint training loss function L_{total}:

$$L_{total} = L_{task} + \alpha \cdot L_{scale} + l_{seg} \cdot \beta \cdot L_{consist} \quad (5)$$

In the formula, L_{task} is the basic task loss (including classification loss and distributed focus loss DFL); α and β is the balanced hyperparameter; l_{seg} is the task indication function (when performing segmentation task training $l_{seg} = 1$, when performing pure detection tasks $l_{seg} = 0$). In order to solve the problem that small targets contribute too little in the global loss calculation and are easily marginalized by the network optimization process, this paper specially designed an adaptive scale-aware loss (Adaptive Scale-aware Loss, L_{scale}). This loss is based on the relative area ratio of the real bounding box in the entire image, and dynamically adjusts the gradient return weight of the bounding box regression:

$$L_{scale} = \frac{1}{N_{pos}} \sum_{i=1}^{N_{pos}} \left(2 - \frac{w_i h_i}{WH}\right) L_{CIoU}\left(B_i, \hat{B}_i\right) \quad (6)$$

in, N_{pos} Represents the number of positive samples in the current batch; w_i and h_i Respectively represent the i The width and height of a real target bounding box; W and H are the global width and height of the input image. B_i and $\hat{B}_i$ They are the real box and the predicted box respectively. This formula constructs a negative correlation weight distribution mechanism: when the target relative area $\frac{w_i h_i}{WH}$ The smaller the target (that is, a tiny target), the closer its loss amplification factor is to 2.0; conversely, the coefficient of a large target approaches 1.0. This dynamic weighting mechanism forces the network to significantly amplify the gradient response of small-scale damage, which greatly enhances the model's upper limit of perception of extremely small-scale targets from the perspective of loss constraints.

3.3.2 Semantic segmentation extension and transfer learning

Compared with simple rectangular box detection, obtaining pixel-level polygon masks such as network cracks, irregular peeling, and gradient rust faces the dual challenges of scarcity of annotation data and difficulty in network convergence. To this end, the DBM architecture proposes a cross-task transfer learning strategy of "detection-guided segmentation optimization"[55]. Since the DBM detection model has learned extremely robust damage boundary priors and feature representations from massive data, when we train complex segmentation branches such as DBM-Conseg and DBM-Stlseg, we do not initialize them from scratch, but directly load the pre-trained backbone and neck weights of the same-level DBM detection model. In order to ensure the smooth transition of feature representation during the transfer process and prevent catastrophic forgetting, this paper introduces a cross-task multi-scale consistency loss ($L_{consist}$）：

$$L_{consist} = \sum_{j\in\{P2,P3,P4,P5\}} MSE\left(F_{seg}^{j}, F_{det\_pre}^{j}\right) \tag{7}$$

In the formula, MSE Represents the mean square error calculation; j Traverse the pyramid levels; F_{seg}^{j} For the current segmentation network in the j The activation feature map of the layer, and $F_{det\_pre}^{j}$ It is the prior feature output of the pre-trained frozen detection network at the same layer. This loss forces the feature encoding space of the segmentation network to be aligned with the mature detection feature space, enabling detection-guided segmentation training. Engineering practice shows that this integrated detection and segmentation strategy not only greatly shortens the convergence period of the segmentation network, but also greatly improves the network's pixel-level resolution accuracy for complex edges of diseases under severely degraded backgrounds.

4. Experimental setup

In order to comprehensively verify the effectiveness of the DBM unified visual perception architecture and cross-task transfer learning strategy proposed in this article in complex bridge inspection scenarios, this section describes the software and hardware environment, implementation details, hyperparameter configuration of model training, and standard evaluation indicators used to measure model performance. These experimental settings provide a standardized benchmark reference for subsequent algorithm evaluation based on the DBM-65k dataset.

4.1 Implementation details and training strategies

Due to the huge amount of data in DBM-65k, it places extremely high demands on the computing power and memory capacity of computing resources. All training tasks in this study were completed on a high-performance workstation equipped with NVIDIA RTX Pro 6000 (48GB). The deep learning framework is based on PyTorch 2.1 and Ultralytics 8.1 series.

During training, the model input resolution was set to 640 × 640 pixels. The AdamW optimizer with momentum (Momentum=0.937) is used, the initial learning rate is set to 0.001, and the weight attenuation is 0.0005. For large-scale data, the Cosine Annealing scheduling algorithm is used to train for 500 Epochs, and a warm-up period of 20 Epochs is set. For the segmentation task (DBM-Conseg/Stlseg), this paper adopts a transfer learning strategy based on detection prior: first train the DBM base model on the full detection subset, and then transfer the learned deep feature weights to the segmentation network for fine-tuning.

Table 4 Experimental environment and key hyperparameter configuration

category	Parameter item	configuration value
Hardware environment	GPU	NVIDIA RTX Pro 6000 (48GB)
	CPU	Intel Xeon Gold 6330 @ 2.00GHz
software environment	CUDA	CUDA 12.1
	Framework	PyTorch 2.1 / Ultralytics
hyperparameters	Input Size	640×640
	Batch Size	32
	Learning Rate	**0.001 (Initial)**
	Optimizer	AdamW

4.2 Joint loss parameter setting

The joint loss function proposed in this article L_{total}Optimization direction for balancing tasks at different levels. The equilibrium coefficient in equation (1) α andβIt is crucial to the performance of the model in extreme multi-scale scenarios.

Based on multiple rounds of hyperparameter search experiments, this paper weights the scale-aware lossαSet to 1.25. Compared with the default weight, slightly increase itαHelps force the network to pay close attention to the tiny building blocks in DBM-Comp that occupy very few pixels early in training. Cross-task consistency weightβIt is set to 0.15 to provide moderate feature constraints to ensure that the segmentation branch does not deviate from the strong physical boundaries established by the detection branch when learning the damage topology.l_{seg}As a Boolean indicator, it is set to 0 in the pure detection training phase and set to 1 in the detection and segmentation integration phase to achieve dynamic switching between tasks.

4.3 Evaluation indicators

In order to comprehensively and objectively evaluate the performance of the DBM model in detection and segmentation tasks, this article uses standard evaluation indicators in the field of computer vision, including precision (Precision, P), recall (Recall, R), and mean average precision (mAP).

$$P = \frac{TP}{TP + FP}, R = \frac{TP}{TP + FN} \tag{8}$$

Among them, TP (True Positives) represents the number of correctly detected diseases; FP (False Positives) represents the number of false detections; FN (False Negatives) represents the number of missed detections.

Mean average precision (mAP) is the core indicator to measure the comprehensive performance of the model. This article mainly reports mAP@0.50 and mAP@0.50:0.95.

$$mAP = \frac{1}{C}\sum_{i=1}^{C} AP_i, AP = \int_0^1 P(R)dR \tag{9}$$

where C represents the total number of categories in the sub-dataset. For segmentation tasks, this paper also reports Mask mAP based on pixel-level mask calculation. In addition, multi-scale evaluation benchmarks are defined based on the ratio of the target bounding box area to the total area of the entire image: mAPs (tiny targets with an area ratio < 1%), mAPm (medium targets with an area ratio between 1% and 5%), and mAPl (macro targets with an area ratio > 5%). This evaluation mechanism based on relative scale division can more accurately and quantitatively verify the perceptual gain of the model for multi-scale targets.

5. Analysis of experimental results

This chapter introduces the comprehensive verification of the effectiveness of the DBM unified architecture in complex bridge inspection tasks through quantitative indicator evaluation and qualitative visual analysis. The overall performance difference between the DBM series models and the baseline network YOLOv12 was compared, and the specific performance of the model under different extreme scales, multiple damage categories, and complex environmental noise was discussed.

5.1 Evaluation Curve and Training Analysis

In order to carefully evaluate the feature recognition and robustness of the DBM model when dealing with bridge damage with different physical attributes, this section analyzes in detail the precision-recall (P-R) curve characteristics of the model in detection and segmentation tasks, the distribution rules of the classification confusion matrix, and the convergence dynamics of the loss function under the cross-task transfer learning strategy.

5.1.1 Macro and micro object detection analysis

In the macro and micro object detection tasks, the P-R curve and confusion matrix presented in Figure 9 reflect the difficulty of bridge disease identification and the response performance of the DBM model. In the concrete subject detection task DBM-Con, the overall mAP@0.50 of the model reached 0.568. Although cracks (Crack) and peeling (Peeling) achieved average accuracy of 0.651 and 0.601 respectively, the confusion matrix revealed a phenomenon: the model has a high degree of mutual misjudgment with the background (Background) in these categories. The proportion of real background misidentified as diseases is distributed between 0.15 and 0.41, while the proportion of real diseases missed and identified as background also reaches 0.33 (cracks) and 0.47 (surface rust and holes). This is because the formwork gaps, dry shrinkage textures, water-stained patches on the real concrete surface and real fine cracks or early surface rust have extremely high visual isomorphism in terms of pixel-level textures. In contrast, in the steel structure main body detection task (DBM-Stl), where metal diseases have higher optical contrast, the model shows higher performance. Its overall mAP@0.50 is as high as 0.930, in which the P-R curve of early coating failure (Failure) and peeling (Peeling) almost fits the top of the coordinate axis, and the diagonal value of the confusion matrix is stable above 0.94, proving that the model is accurate in extracting the apparent degradation characteristics of steel structures.

However, in the detailed micro-component inspection task (DBM-Comp), the performance distribution shows significant polarization. The model achieves high-precision capture of normal bolts (Normal) and cable damage (Broken-cab) (mAP is 0.932 and 0.791 respectively). However, the confusion matrix pointed out that for the small defect of loose bolts (Loose), the model had a serious classification bottleneck. The loose bolts were misjudged by the network as normal bolts (0.35) or rusty bolts (0.65). On the one hand, the visual characteristics of a loose bolt are only the relative displacement of a few pixels between the nut and the washer, which can easily be smoothed out by downsampling operations in long-distance drone photography. On the other hand, once the surface of a loose bolt is accompanied by rust, the large-area rust texture features will cover up the weak geometric deformation features in the forward propagation of the convolution layer, causing the network to shift the feature focus. Although this lowers the index of this subcategory, it also truly reflects the physical limits of current reliance on pure 2D vision for state assessment of extremely small structures, highlighting the need to build DBM-65k, a complex multi-state benchmark.

DBM-Con DBM-Stl DBM-Comp

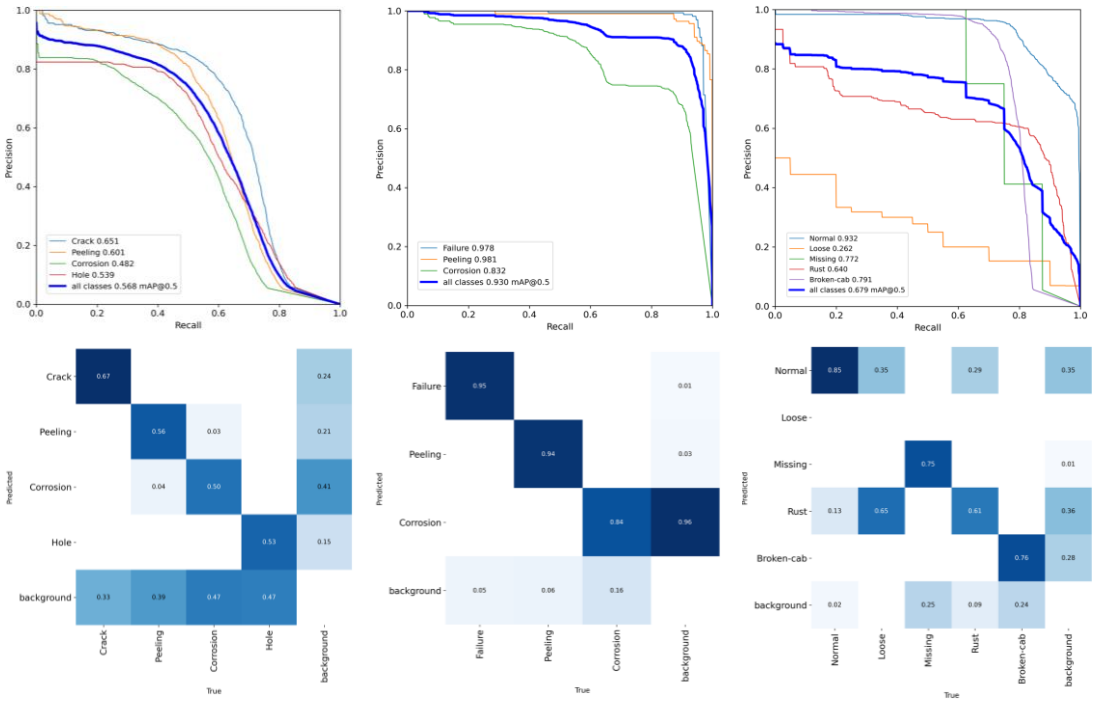

Figure 9. P-R curve and confusion matrix of detection task.

5.1.2 Semantic segmentation analysis of complex damage

For more demanding pixel-level segmentation tasks, the DBM-seg model also demonstrates feature analysis capabilities. In the DBM-Conseg task, the model segmented surface spalling and exposed rebar with extremely high accuracy, and the P-R curve covered a large area. The confusion matrix further confirmed that the model can accurately peel off these two diseases with strong geometric depth with 92% and 82% confidence. However, it is worth noting that the mask accuracy of the efflorescence and water damage categories is relatively low (0.534), and nearly 49% of the pixels in the confusion matrix are judged as background by the network. From a pathological and optical perspective, water stains and efflorescence often present highly divergent, color-gradient topological features on the concrete surface, lacking clear physical change boundaries similar to cracks or spalling. This semantic boundary ambiguity makes it difficult for the network to establish an absolute segmentation threshold when performing pixel-level discrimination, resulting in a conservative trend in mask edge prediction.

In the DBM-Stlseg steel structure corrosion segmentation, the model maintained a high-precision segmentation of more than 0.92 for different deterioration levels (general, poor, severe), and the diagonal elements of the confusion matrix were highly concentrated, proving the dynamic alignment module (DAF)'s strong ability to fit irregular corrosion boundaries. In addition, the dynamic curves of the loss functions (Box Loss and Seg Loss) on the training and validation sets show that thanks to the "detection prior-guided cross-task transfer learning" strategy proposed in this article, the segmentation network achieved rapid gradient descent in the early stages of training, and the validation set curves converged smoothly without obvious overfitting oscillations. This fully demonstrates that transferring the spatial prior knowledge of mature detection networks to segmentation tasks can greatly accelerate the convergence process of complex edge parsing.

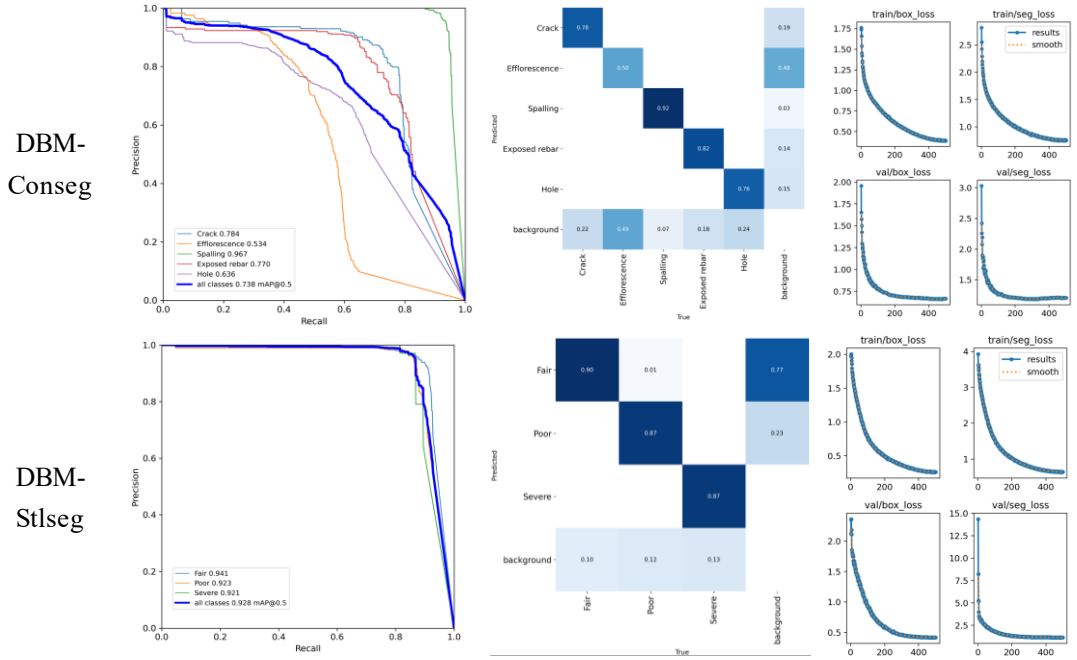

Figure 10. P-R curve, loss curve and confusion matrix of segmentation task.

5.2 Quantitative evaluation of performance of DBM model

After analyzing the internal convergence and category identification mechanism of the model, this section further reveals the specific performance gains of the DBM unified architecture in different evaluation dimensions and extreme scales through a comprehensive and rigorous quantitative comparison with the advanced baseline model YOLOv12.

5.2.1 Evaluation index performance comparison

Based on the test results of the five core sub-datasets, the DBM series models have achieved improvements over the baseline model YOLOv12 in key recall rates (Recall), comprehensive average precision (mAP@0.50), and strict threshold indicators reflecting pixel-level high-precision positioning (mAP@0.50:0.95). The experimental data shows a characteristic: the YOLOv12 model shows a weak accuracy advantage in some tasks (such as DBM-Conseg's Box Precision reaching 0.8542, DBM-Stl's Precision reaching 0.8627). However, this precision lead comes at the expense of a large amount of recall. In a real engineering scenario, this means that the baseline model adopts an extremely conservative prediction strategy and responds only to those macroscopic diseases with extremely obvious characteristics and occupying a large area of pixels, thereby completely ignoring a large number of difficult diseases with blurred edges and small scales. This is unacceptable in bridge inspections where prevention of missed inspections is the first priority.

In contrast, the DBM model actively lowers the detection limit of small blur defects after introducing a dynamic small target head and scale-aware loss. Although this introduces a few background false positives, the overall recall rate is improved across the board. On the Mask mAP@0.50:0.95 indicator that measures the absolute fit of pixel-level boundaries, DBM-seg achieved excellent results of 0.4728 and 0.7347 in the concrete and steel structure segmentation tasks, respectively. This hard-core performance of greatly improving mask accuracy under strict thresholds at the expense of very small accuracy fluctuations strongly proves that the DBM architecture has more refined and robust pixel-level perception when dealing with irregular topological diseases.

Table 5 DBM family macro and micro task quantitative performance evaluation indicators

Dataset	Model	Precision	Recall	mAP@0.50	mAP@0.50-0.95
DBM-Con	DBM	0.6681	0.5562	0.5683	0.3271
	YOLOv12	0.6724	0.5489	0.5615	0.3182
DBM-Stl	DBM	0.8690	0.9159	0.9301	0.8050
	YOLOv12	0.8627	0.9016	0.9143	0.7828
DBM-Comp	DBM	0.5293	0.7358	0.6794	0.4388
	YOLOv12	0.5360	0.7185	0.6657	0.4219

Dataset	Model	Metric Type	Precision	Recall	mAP@0.50	mAP@0.50-0.95
DBM-Conseg	DBM-seg	Box	0.8514	0.7438	0.8115	0.6639
		Mask	0.7932	0.6949	0.7381	0.4728
	YOLOv12-seg	Box	0.8542	0.7316	0.8037	0.6514
		Mask	0.7850	0.6813	0.7245	0.4586
DBM-Stlseg	DBM-seg	Box	0.9674	0.8747	0.9333	0.8671
		Mask	0.9540	0.8704	0.9283	0.7347
	YOLOv12-seg	Box	0.9585	0.8649	0.9212	0.8503
		Mask	0.9601	0.8558	0.9150	0.7137

5.2.2 Analysis of subcategories of each dataset

Targeted analysis of the quantitative performance of the DBM model in respective sub-task subcategories can provide a clearer understanding of its capability boundaries, as shown in Table 6. In DBM-Con, the model has extremely high positioning accuracy (0.7102) for holes with clear edges, but is limited by the change in shadow perspective caused by the depth of the disease, and its recall rate is low. In DBM-Stl, the model shows strong capabilities, especially for coating failure (Failure), the comprehensive mAP@0.50:0.95 is as high as 0.8902, indicating that the model can accurately quantify the subtle thickness difference in the early stage of coating peeling.

In the challenging DBM-Comp task, in addition to the loose category, the model's recall rate for missing bolts (Missing) reached 0.7500. Considering that a missing bolt hole only appears as a few dark pixels in a long-distance perspective, this metric fully proves that the network still maintains an effective spatial response in areas with extremely poor features. In the segmentation fields of DBM-Conseg and DBM-Stlseg, the model not only maintains a high accuracy of over 0.87 in common damage types such as cracks and spalling, but also achieves mask accuracies of 0.6723, 0.7229 and 0.8089 respectively in the highly subjective steel structure corrosion rating, perfectly realizing the leap from single visual inspection to engineering disease grading and quantification.

Table 6 Performance summary of each individual subcategory of the DBM family

Dataset	Class	Precision (B)	Recall (B)	mAP@50 (B)	mAP@50-95 (B)
DBM-Con	Total	0.6681	0.5562	0.5683	0.3271
	Crack	0.6846	0.6520	0.6509	0.4595
	Peeling	0.6783	0.5773	0.6013	0.3951
	Corrosion	0.5995	0.4979	0.4823	0.2469
	Hole	0.7102	0.4975	0.5386	0.2071
DBM-Stl	Total	0.8690	0.9159	0.9301	0.8050
	Failure	0.9658	0.9527	0.9780	0.8902
	Peeling	0.9062	0.9492	0.9807	0.8291
	Corrosion	0.7352	0.8460	0.8317	0.6956

DBM-Comp	Total	0.5293	0.7358	0.6794	0.4388
	Normal	0.6890	0.9766	0.9325	0.6183
	Loose	0.2918	0.3000	0.2615	0.1485
	Missing	0.3975	0.7500	0.7720	0.4833
	Rust	0.5168	0.8784	0.6405	0.4311
	Broken-cab	0.7514	0.7741	0.7907	0.5126
Dataset	Class	Precision (M)	Recall (M)	mAP@50 (M)	mAP@50-95 (M)
DBM-ConSeg	Total	0.7932	0.6949	0.7381	0.4728
	Crack	0.7962	0.7801	0.7838	0.462
	Efflorescence	0.7443	0.4819	0.5344	0.2757
	Spalling	0.9673	0.9209	0.9668	0.878
	Exposed rebar	0.7383	0.746	0.77	0.5101
	Hole	0.7199	0.5458	0.6356	0.2381
DBM-Stlseg	Total	0.954	0.8704	0.9283	0.7347
	Fair	0.9466	0.8842	0.9412	0.6723
	Poor	0.9578	0.8586	0.9228	0.7229
	Severe	0.9577	0.8684	0.9208	0.8089

5.2.3 Analysis of performance gain of extreme scale targets

In order to further strip away the dominant and concealing effect of large targets on the overall evaluation indicators, this section specifically extracts the hierarchical accuracy (Small, Medium, Large) at multiple scales for in-depth analysis. In YOLOv12, the model's detection of large targets (mAPl) has tended to be saturated or even slightly dominant, but serious feature faults and performance degradation have occurred when facing small defects (mAPs). After the introduction of the DBM architecture, the model's detection focus accurately shifted to small-scale damage while maintaining a steady increase in overall accuracy.

Specifically, in the three major tasks of DBM-Con, DBM-Stl and DBM-Comp, the DBM model achieved absolute gains of up to +3.35%, +5.37% and +4.28% in small target accuracy (mAPs@0.50) at the expense of only a weak decrease in large target accuracy (mAPl). This strategy is in line with the underlying logic of preventing missed detections in drone bridge inspections. Macroscopic large-area diseases are more difficult to miss, while hidden micro-defects at long distances are a problem with traditional algorithms. The data in Table 7 rigorously proves that the dynamic small object detection head (Dynamic P2 Head) successfully explores the resolution granularity of the feature pyramid, and the dynamic alignment fusion module (DAF) and scale-aware loss jointly work together to effectively amplify the gradient return weight of small-scale damage and break through the perception limitations of remote micro-defect troubleshooting.

Table 7 Extreme scale target hierarchical performance gain table

Dataset	Model	mAP@0.50	mAP_s@0.50	mAP_m@0.50	mAP_l@0.50	mAP_{s_gain}
DBM-Con	DBM	0.5683	0.3847	0.6183	0.7654	3.35%
	YOLOv12	0.5615	0.3512	0.6148	0.7725	—
DBM-Stl	DBM	0.9301	0.8461	0.9483	0.9829	5.37%
	YOLOv12	0.9143	0.7924	0.9416	0.9852	—
DBM-Comp	DBM	0.6794	0.4856	0.7251	0.8587	4.28%
	YOLOv12	0.6657	0.4428	0.7215	0.8643	—

5.3 Qualitative visualization in complex scenarios

This section intuitively demonstrates the engineering reliability and visual interpretability of the DBM model under complex illumination and background, as well as distance and distance, through direct prediction frame and visualization comparison.

5.3.1 Investigation and comparison of long-distance small targets

In the qualitative comparison of long-distance drone cruise perspectives (as shown in Figure 11), the DBM model demonstrated detailed feature capture capabilities and identification robustness that significantly exceeded the baseline network. In the DBM-Comp micro-component inspection scenario, when faced with a bolt group matrix that is densely arranged and has a very low pixel ratio in the entire image, the baseline model YOLOv12 uses the traditional continuous downsampling mechanism, resulting in the irreversible loss of spatial geometric information in the deep feature map. In the visualization results, it is manifested as a large area of missing prediction boxes and missed detection, and it completely loses the ability to identify the fine-grained identification of the specific service status of the bolts.

In contrast, the DBM model relies on the P2 dynamic detection head to retain shallow high-resolution features, and is supplemented by a scale and space dual perception attention mechanism. It not only successfully achieves accurate anchoring of the spatial position of each independent tiny bolt in a wide-angle and large field of view, constructing a dense and complete prediction frame network, but also accurately identifies individual abnormal nodes that are in a rusted or missing state in the complex steel truss background, and its decision boundary shows extremely sharpness. In addition, in the long-distance inspection of a subset of DBM-Stl steel structures, the DBM model also demonstrated the sensitivity of microscopic feature extraction, successfully capturing early coating failure areas with extremely weak edge gradients and extremely low visual contrast, while the baseline model completely ignored these early signs of deterioration due to insufficient feature response. In long-distance cable inspection, DBM can also effectively capture the damaged area of the cable. This series of qualitative visual results is highly consistent with the small target hierarchical performance index (mAPs@0.50) shown previously, which strongly proves the absolute technical advantages and engineering practical value of this customized architecture in solving the bottleneck of extremely small target perception.

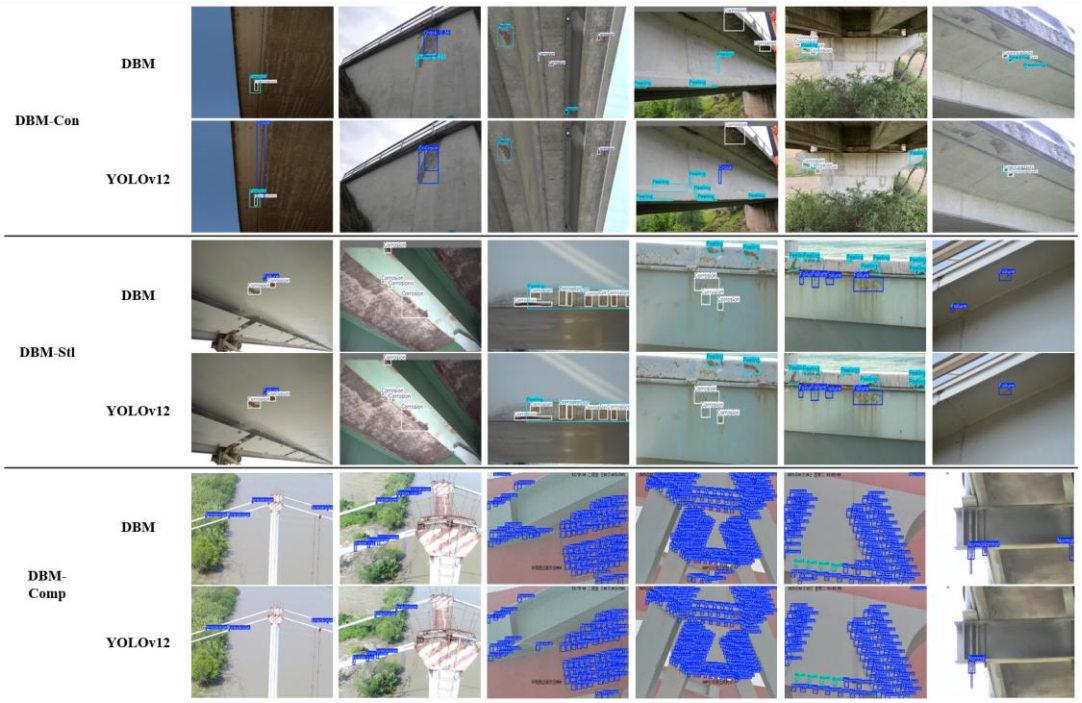

Figure 11. Visual comparison of scenarios with minute defects at long distances.

5.3.2 Close range, large field of view and complex background analysis

Real bridge inspection images are generally accompanied by complex optical distortion and non-structural artifacts. As shown in Figure 12, the model's generalization ability was evaluated under interference conditions such as strong backlight shadows, dynamic projections from drones, and surface water stains. The visual comparison results highlight the flaws of the baseline network in handling non-semantic interference. YOLOv12 is susceptible to interference factors such as linear water flow traces on the concrete surface or rotor projection, resulting in a large number of high-confidence false positive predictions. This phenomenon stems from the over-sensitivity of conventional convolutional networks to high-frequency edge features, which fails to effectively decouple real physical diseases and false optical shadows.

The DBM model demonstrates background suppression and semantic decoupling capabilities. Thanks to the network's adaptive perception of spatial geometric deformation in the multi-scale fusion stage, DBM successfully peels off environmental artifacts that have nothing to do with real damage. Even in harsh lighting areas where disease features such as initial rust spots or shallow peeling are in extremely low visual contrast with healthy substrates, the model can still accurately anchor the true physical boundaries of the damage without drifting or false alarms in the prediction frame. This set of visual evidence conclusively shows that the DBM architecture has significant advantages in resisting complex natural noise and maintaining robustness in open environments.

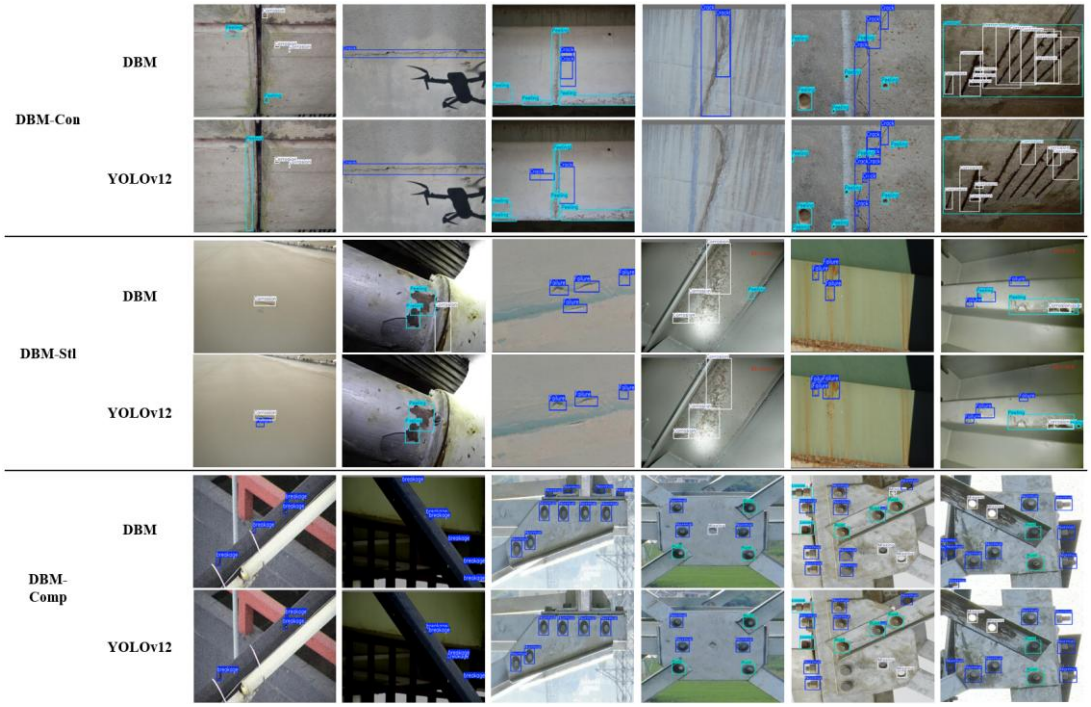

Figure 12. Visualized comparison of detection under complex backgrounds.

5.3.3 Comparative analysis of fine segmentation

In the visual analysis of pixel-level segmentation results (as shown in Figure 13), this article simultaneously displays the colored polygon mask output by DBM-seg and its corresponding feature attention heat map. Observing the high-response area of the heat map, we can find that the feature activation of DBM-seg is highly focused on the core texture and edge mutations of the disease. When dealing with concrete cracks that present an irregular network of intersections, or severe corrosion of steel structures whose boundaries exhibit a highly smudged gradient state, the mask boundaries generated by DBM-seg are smoother, more continuous, and highly consistent with the real physical topology than the baseline model; while the masks of the baseline model frequently have jagged edges, internal voids, or topological fractures. DBM-seg's high-precision ability to depict complex disease geometric shapes not only eliminates mask faults in visual representation, but also provides a highly reliable underlying pixel basis for subsequent engineering applications such as automatic calculation of the actual surface area of the disease and assessment of structural bearing capacity degradation and other quantitative decision-making processes.

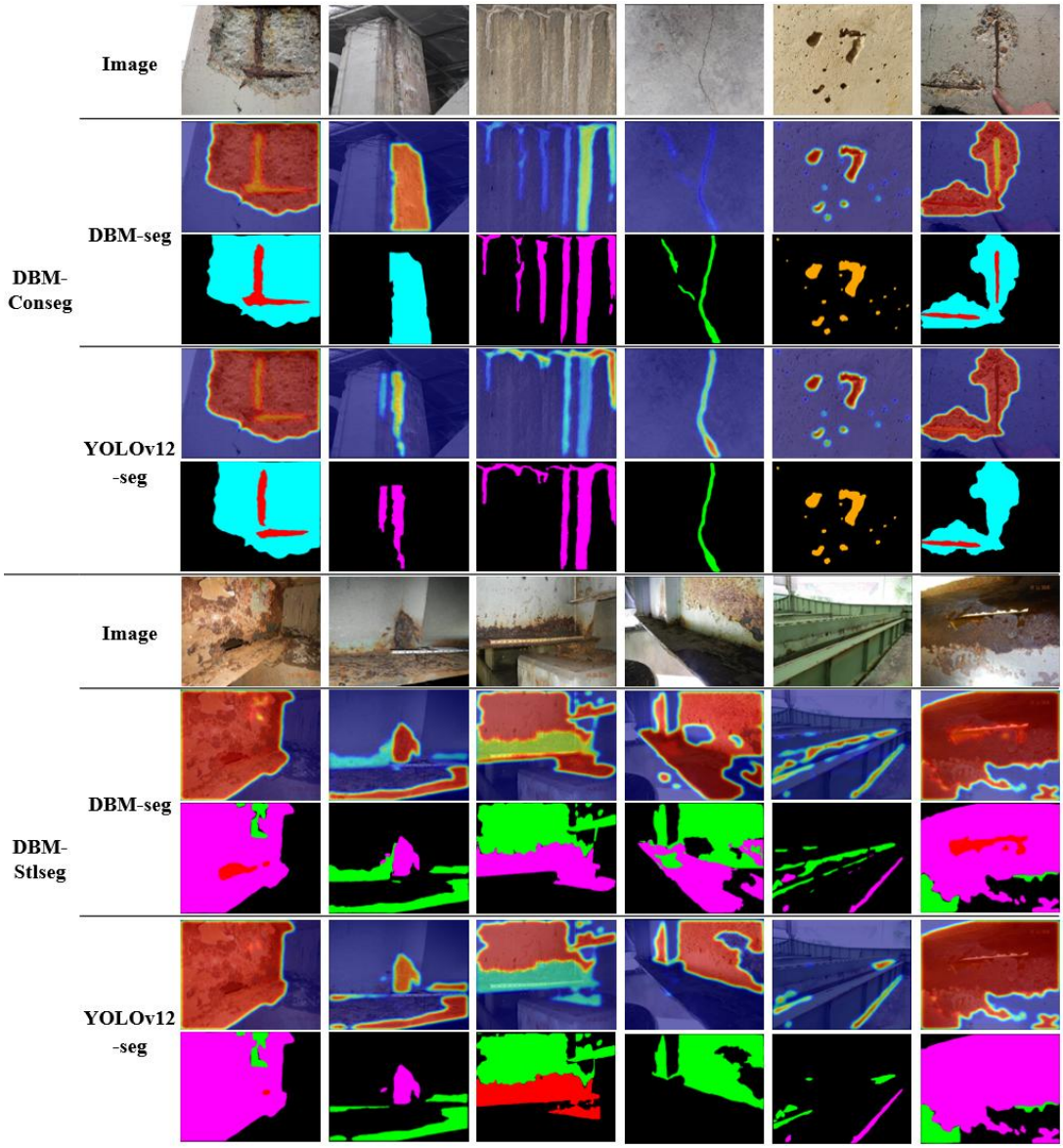

Figure 13. Pixel-level segmentation results of complex structures.

5.4 Ablation and comparison experiments

In order to further verify the effectiveness of the DBM model in the task of detecting steel structure bridge damage, this section conducts comparative experiments and ablation experiments on the DBM-Stl subset. The comparative experiment selects representative advanced models in the field of general object detection in recent years. Among them, Gold-YOLO, YOLOv9, YOLOv10, YOLO11 and YOLOv12 represent the development route of YOLO series detectors in recent years, while RT-DETRv2 and D-FINE represent real-time DETR detection frameworks. All models were compared using the same data partitioning, input resolution, and evaluation metrics.

Table 8 Comparison experiment on DBM-Stl

Model	Precision	Recall	mAP@0.50	mAP@0.50
Gold-YOLO	0.8589	0.8952	0.9058	0.7746
YOLOv9	0.8615	0.8998	0.9108	0.7796

YOLOv10	0.8642	0.9041	0.9126	0.7813
RT-DETRv2	0.8604	0.9047	0.9122	0.7864
YOLO11	0.8661	0.9088	0.9175	0.7896
D-FINE	0.8674	0.9107	0.9218	0.7969
YOLOv12	0.8627	0.9016	0.9143	0.7828
DBM-Stl	0.8690	0.9159	0.9301	0.8050

As can be seen from Table 8, there are obvious differences in the performance of different general detection models on DBM-Stl. Gold-YOLO's mAP@0.50 is 0.9058, and mAP@0.50-0.95 is 0.7746, indicating that its general feature aggregation ability can complete steel structure damage detection, but it still has shortcomings in complex corrosion and coating boundary positioning. The mAP@0.50 of YOLOv9 and YOLOv10 increased to 0.9108 and 0.9126 respectively, which are overall better than Gold-YOLO, indicating that the new generation YOLO structure has better adaptability in multi-scale feature expression. The mAP@0.50 of RT-DETRv2 is 0.9122, slightly lower than YOLOv10, but mAP@0.50-0.95 reaches 0.7864, which is higher than 0.7813 of YOLOv10, indicating that the DETR class model has certain advantages in high threshold positioning.

YOLO11 and D-FINE further improved the detection performance. YOLO11's mAP@0.50 and mAP@0.50-0.95 were 0.9175 and 0.7896 respectively, while D-FINE reached 0.9218 and 0.7969, which is the method closest to DBM-Stl in the comparison model. In contrast, YOLOv12, as the baseline model in this article, achieved mAP@0.50 of 0.9143 and mAP@0.50-0.95 of 0.7828. It has strong basic detection capabilities, but its Recall is 0.9016, which is lower than YOLO11, D-FINE and DBM-Stl. DBM-Stl achieved the highest results in all indicators, with Precision, Recall, mAP@0.50 and mAP@0.50-0.95 being 0.8690, 0.9159, 0.9301 and 0.8050 respectively, indicating that dynamic alignment fusion and high-resolution detection head design for multi-scale diseases of steel structure bridges can further improve the damage detection rate and positioning quality under strict thresholds.

Table 9 Ablation experiment on DBM-Stl

experimental group	DAF	P2 Head	Precision	Recall	mAP@0.50	mAP@0.50-0.95
Baseline (YOLOv12)			0.8627	0.9016	0.9143	0.7828
+ DAF	✓		0.8655	0.9085	0.9225	0.7965
+ P2 Head		✓	0.8640	0.9102	0.9208	0.7912
Full DBM-Stl	✓	✓	0.8690	0.9159	0.9301	0.8050

Ablation experiments further illustrate the contribution of the two core modules in DBM. Taking YOLOv12 as the baseline, the model's mAP@0.50 is 0.9143, and mAP@0.50-0.95 is 0.7828. After introducing DAF alone, mAP@0.50 increased to 0.9225, and mAP@0.50-0.95 increased to 0.7965, indicating that dynamic alignment fusion can alleviate the spatial misalignment problem in cross-layer feature fusion and improve the quality of steel structure damage boundary positioning. After introducing Dynamic P2 Head alone, Recall increased from 0.9016 to 0.9102, and mAP@0.50 increased to 0.9208, indicating that the high-resolution detection layer has a complementary role in small-scale coating failure

and local corrosion. When DAF is used with P2 Head at the same time, the complete DBM-Stl achieves the highest value in all four indicators, indicating that the two have complementary effects in multi-scale feature alignment and small target response enhancement.

5.5 Testing in real bridge environment

After completing the offline benchmark evaluation, scale hierarchical analysis and ablation experiment, this paper further selected the Fourth Nanjing Yangtze River Bridge as the real bridge environment test object to test the operational feasibility and generalization ability of the DBM architecture in open engineering scenarios. The Fourth Nanjing Yangtze River Bridge is an important cross-river channel in Nanjing's cross-river transportation system. The bridge site spans the main channel of the Yangtze River. The bridge structure has a large scale, multiple types of components, and a complex service environment. Its main structure includes multiple types of structural units such as the main tower, main cables, slings, steel box girders, bridge deck ancillary components and connecting nodes. It can simultaneously present multiple types of detection objects such as concrete apparent degradation, steel structure corrosion, cable anomalies and bolt connection status. Affected by the hot and humid environment on the river surface, wind load, traffic load and long-term service, this type of long-span bridge is prone to problems such as coating aging, local corrosion, state changes of connecting components, and complex lighting background interference during actual inspections. Therefore, it is suitable as an on-site scenario to verify the engineering applicability of the UAV intelligent detection model.

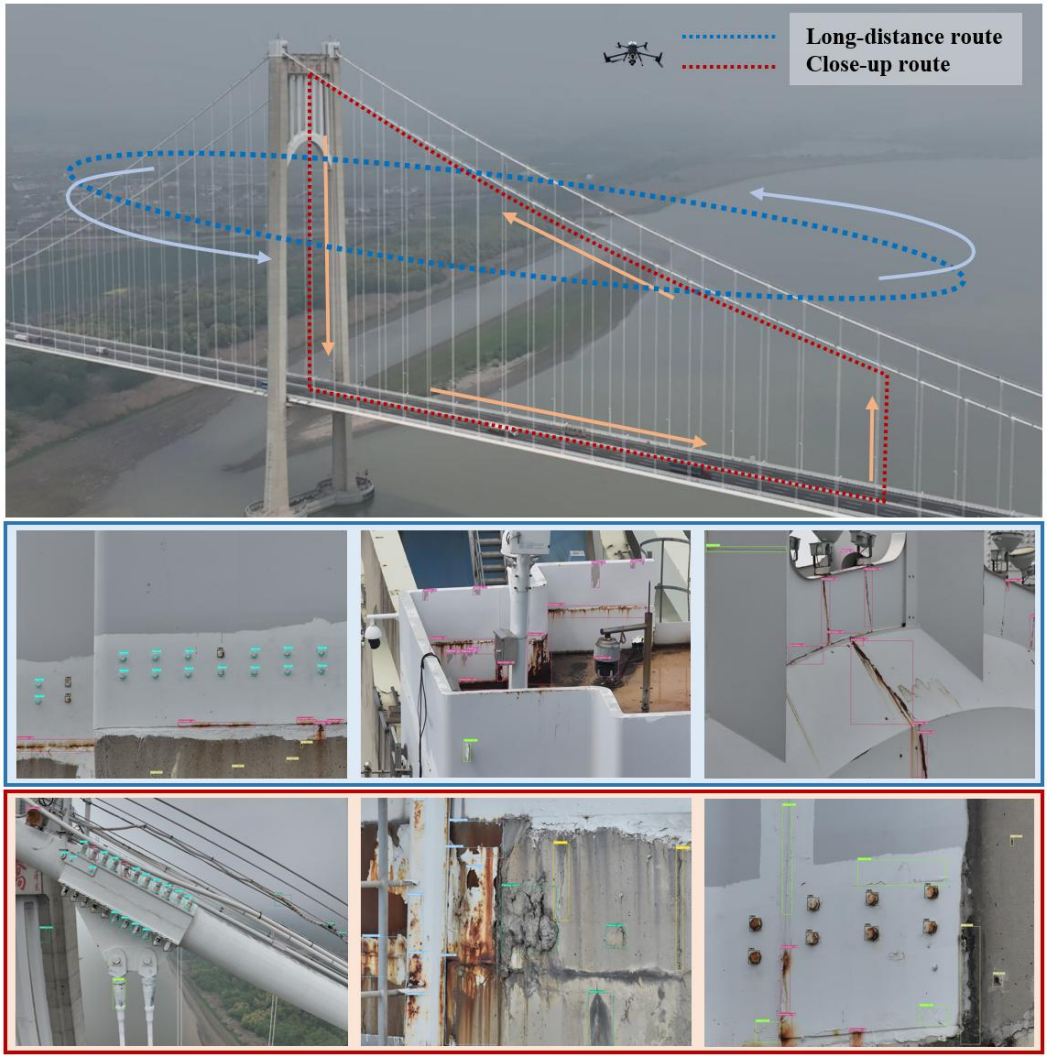

Figure 14. Real-world disease detection based on the DBM framework.

The on-site test process relies on the intelligent inspection pipeline shown in Figure 14. The drone adopts a flight method that combines long-distance structural cruise and short-distance local review: long-distance cruise is mainly used to quickly cover the main tower, cables, beam outer surface and large-scale steel structure areas to screen obvious apparent degradation and suspected damage locations; short-distance review is to supplement the collection of connecting nodes, bolt groups, local corrosion areas and coating abnormal areas to improve the visibility of small components and fine-grained diseases. This testing method can simultaneously cover large-scale structural apparent diseases and long-distance small component defects, and is consistent with the data organization logic of "main structure apparent inspection + key component fine inspection" in DBM-65k.

The test results show that DBM can better continue the multi-scale detection capabilities in offline experiments in the live images of the Fourth Nanjing Yangtze River Bridge. Facing multiple structural backgrounds such as main towers, steel beams, cables, and connecting plates, the model can locate large-scale apparent degradation areas on the main bridge structure, and conduct status inspections of bolt groups, local cable anomalies, and steel structure corrosion areas from a long-distance perspective. Especially in images with overlapping complex components, dense background textures, and obvious target scale differences, DBM can still maintain a relatively concentrated detection response, indicating that the multi-type and multi-scale training distribution constructed in DBM-65k has a certain supporting

effect on on-site generalization.

At the same time, the real bridge environment also exposes sources of interference that are difficult to fully cover with offline data. Affected by the reflection of the river surface, strong light changes, environmental smearing, changes in the drone's perspective and the appearance of unseen components, the model may still misjudge local rusty cables as damaged, misjudge the cylindrical camera on the surface of the cable tower as a rusted area, or misjudge healthy geometric components as early coating failure under extreme vertical viewing angles. These misidentifications mainly come from the local similarity between the target appearance and the disease characteristics, rather than the complete loss of the model's positioning ability. Overall, the on-site test of the Fourth Nanjing Yangtze River Bridge shows that DBM has application potential for real bridge inspection processes, but its on-site deployment still needs to be combined with manual review, confidence threshold screening and subsequent multi-modal information supplementation to further improve detection reliability in an open environment.

6. Conclusion and future prospects

This paper proposes a complete intelligent sensing solution from the perspective of data-driven and architecture customization, aiming at the core pain points such as scale differences, complex damage types, and limited scale of existing datasets faced in UAV bridge inspections. By constructing the ultra-large-scale benchmark dataset DBM-65k and developing the DBM unified architecture model, this study established a new paradigm of inspection and segmentation integration through a macro-micro structure hierarchical system, taking a key step for UAV inspection of bridge structures from laboratory algorithms to large-scale engineering implementation.

The research results of this article demonstrate the great value of deep collaboration between large-scale real-world data and targeted network architecture. The core contributions are summarized as follows:

1. A large-scale, multi-damage type "macro-micro layered" bridge benchmark database DBM-65k was constructed. The dataset contains more than 65,000 real inspection images and nearly 300,000 annotated instances, covering five core task streams such as concrete, steel structures, cables and bolts, breaking the isolation of existing datasets in terms of material and scale.

2. Developed a series of DBM models and established a unified multi-task assessment framework. By deploying different detection and segmentation branches on a unified feature extraction layer, DBM achieves simultaneous perception of macroscopic subject damage and microscopic key component status, significantly improving the model's generalization ability in heterogeneous inspection environments.

3. Proposed the dynamic alignment fusion module (DAF) and dynamic small object detection head (DynamicP2Head). By introducing spatial adaptive alignment of deformation convolution, the collapse problem of small disease features due to downsampling in deep networks was successfully solved, resulting in a maximum absolute gain of 5.37% in small object detection accuracy (mAPs).

4. The adaptive scale-aware loss (AdaptiveScale-awareLoss) is designed to significantly amplify the gradient response of small targets. This loss mechanism enables the network to actively allocate more optimization weights to pixel sparse areas during the training process through negative correlation weight allocation, overcoming the bottleneck of long-distance micro-defect recognition from the constraint level.

5. A cross-task transfer learning strategy driven by detection priors is adopted to achieve the integration of macro and micro detection. By migrating the spatial weights of mature detection models to the segmentation network and introducing cross-task consistency loss, DBM-seg shows excellent pixel-level resolution accuracy when dealing with complex topological diseases such as network cracks and irregular corrosion.

Although the DBM architecture shows excellent robustness in most scenarios, in-depth analysis of the experimental results can still reveal the limitations of the current technology under extreme conditions. First, for some extremely fine-grained microscopic deformations (such as slight loosening of bolts)[56, 57]), since its judgment features only rely on the relative displacement of several pixels, the current recognition rate of the model (mAP@0.50 is 0.2615) still needs to be improved. Under the influence of long distance, strong light interference or image compression noise, the physical characteristics of such targets are easily annihilated by semantics. Secondly, in the segmentation task of complex damage, due to the lack of clear physical boundaries for diseases such as efflorescence and water seepage, the prediction of the mask edge is somewhat conservative, which reflects the limitations of a single visual modality in processing "semantic fuzzy boundaries"[58, 59]. In addition, the stability of the algorithm is still challenged to a certain extent when faced with extreme environmental occlusion or severe motion blur caused by the rapid flight of the drone.[60].

Future research work will focus on the following directions: First, explore multi-modal feature fusion, and solve the blind spots of single visible light perception under shadows, occlusions and drastic changes in illumination by introducing infrared thermal imaging or depth point cloud data; second, study lightweight and high-performance end-side deployment solutions to enable DBM The model can be seamlessly integrated into the UAV embedded computing platform to achieve real-time damage early warning; third, it is combined with the visual large language model (VLM) to build a bridge intelligent diagnosis system with logical reasoning capabilities, thereby realizing a full-link intelligent transformation from disease identification to automatic generation of inspection reports. This research provides a solid data foundation and algorithm reference for building a "unified" bridge structure health detection system.

CRediT authorship contribution statement

Junwen Zheng: Writing – review & editing, Writing – original draft, Validation, Software, Methodology, Investigation, Formal analysis. Hao Feng: Writing – review & editing, Writing – original draft, Validation, Supervision, Software, Methodology, Investigation, Conceptualization. Jinghuan Zhang: Writing – review & editing, Software, Investigation, Data curation. Jian Zhang: Writing – review & editing, Validation, Software, Investigation.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Model weights have been uploaded and are publicly available at https://bit.ly/4uMvTOa. Other data will be made available on request.

Acknowledgment

The research presented was financially supported by the National Key R&D Program of China (No. 2022YFC3801700), the National Natural Science Foundation of China (No. 52378289), the Research Fund for Advanced Ocean Institute of Southeast University (No. KP202407), the Suzhou Science and Technology Program - Key Core Technologies Project (No. SYG2025117).

Reference

[1] A. Ashmawi, P. Nguyen, A. Jawdhari, State-of-the-art review of machine learning applications for bridge inspections, Advances in Structural Engineering, 29 (2026) 1217-1249.

[2] C. Zhang, Y. Zou, F. Wang, E. del Rey Castillo, J. Dimyadi, L. Chen, Towards fully automated unmanned aerial vehicle-enabled bridge inspection: Where are we at?, Construction and Building

Materials, 347 (2022) 128543.
[3] H. Sun, L. Song, Z. Yu, A deep learning-based bridge damage detection and localization method, Mechanical Systems and Signal Processing, 193 (2023) 110277.
[4] S. Dorafshan, R.J. Thomas, M. Maguire, Comparison of deep convolutional neural networks and edge detectors for image-based crack detection in concrete, Construction and Building Materials, 186 (2018) 1031-1045.
[5] B.F. Spencer Jr, V. Hoskere, Y. Narazaki, Advances in computer vision-based civil infrastructure inspection and monitoring, Engineering, 5 (2019) 199-222.
[6] T. Panigati, M. Zini, D. Striccoli, P.F. Giordano, D. Tonelli, M.P. Limongelli, D. Zonta, Drone-based bridge inspections: Current practices and future directions, Automation in Construction, 173 (2025) 106101.
[7] V. Prakash, C.J. Debono, M.A. Musarat, R.P. Borg, D. Seychell, W. Ding, J. Shu, Structural health monitoring of concrete bridges through artificial intelligence: A narrative review, Applied Sciences, 15 (2025) 4855.
[8] L. Zhou, Y. Jiang, H. Jia, L. Zhang, F. Xu, Y. Tian, Z. Ma, X. Liu, S. Guo, Y. Wu, UAV vision-based crack quantification and visualization of bridges: system design and engineering application, Structural Health Monitoring, 24 (2025) 1083-1100.
[9] G. Xu, Y. Zhang, Q. Yue, X. Liu, A deep learning framework for real-time multi-task recognition and measurement of concrete cracks, Advanced Engineering Informatics, 65 (2025) 103127.
[10] S.S.C. Congress, A.J. Puppala, M.A. Khan, N. Biswas, P. Kumar, Application of unmanned aerial technologies for inspecting pavement and bridge infrastructure assets conditions, Transportation Research Record, 2679 (2025) 529-543.
[11] E. Figueiredo, N. Peres, I. Moldovan, A. Nasr, Impact of climate change on long-term damage detection for structural health monitoring of bridges, Structural Health Monitoring, 24 (2025) 2252-2270.
[12] V.H. Do, T.C. Pham, H.N. Phan, M.H. Nguyen, P.N. Huynh, Deep learning for bridge component classification and damage detection from UAV imagery, Proceedings of the Institution of Civil Engineers-Bridge Engineering, Emerald Publishing Limited, 2025, pp. 1-12.
[13] Z. Yao, Y. Li, H. Fu, J. Tian, Y. Zhou, C.-L. Chin, C.-K. Ma, Research on concrete crack and depression detection method based on multi-level defect fusion segmentation network, Buildings, 15 (2025) 1657.
[14] B. Xu, W. Shao, X. Dong, Drone-based wall crack detection using model-agnostic meta-learning, IEEE Transactions on Automation Science and Engineering, (2025).
[15] L. Liu, H. Gong, Y. Zhou, A. Zhou, L. Cong, A YOLO-based network with dynamic feature pyramid for multi-scale bridge surface defect detection, International Journal of Transportation Science and Technology, (2025).
[16] M.A.-M. Khan, S.-H. Kee, A. Pathan, A.A. Nahid, Image Processing Techniques for Concrete Crack Detection: A Scientometrics Literature Review, Remote. Sens., 15 (2023) 2400.
[17] T. Liu, L. Zhang, G. Zhou, W. Cai, C. Cai, L. Li, BC-DUnet-based segmentation of fine cracks in bridges under a complex background, PLoS ONE, 17 (2022).
[18] J. Zhang, W. Chen, J. Zhang, UAV-based quantitative crack measurement for bridges integrating four-point laser metric calibration and mamba segmentation, Automation in Construction, 182 (2026) 106774.
[19] J. Huang, Y. Zhu, M. Xiong, J.D. Ser, A. Alotaibi, J.P. Papa, K. Muhammad, Efficient bridge damage detection using a lightweight attention-based modeling framework, Computer-Aided Civil and

Infrastructure Engineering, 40 (2025) 4758 - 4773.
[20] X. Zhang, H. Wang, Y.-A. Hsieh, Z. Yang, A. Yezzi, Y.-c. Tsai, Deep Learning for Crack Detection: A Review of Learning Paradigms, Generalizability, and Datasets, ArXiv, abs/2508.10256 (2025).
[21] S. Li, H. Li, W. Lu, Z. Zhou, Combining motion blur removal with intelligent methods for tunnel defect detection, Automation in Construction, 178 (2025) 106394.
[22] M. Maguire, S. Dorafshan, R.J. Thomas, SDNET2018: A concrete crack image dataset for machine learning applications, 2018.
[23] R.-s. Ji, Y. Xu, X. Wang, L. Zhuang, X. Zhang, X. Tang, J. Shi, LBSD-YOLO: A Lightweight YOLOv10-Based Network with Multi-Attention Enhancement for Bridge Surface Defect Detection, ICCK Transactions on Sensing, Communication, and Control, (2026).
[24] Y. Gao, H. Li, W. Fu, Few-shot learning for image-based bridge damage detection, Eng. Appl. Artif. Intell., 126 (2023) 107078.
[25] D. Kumar, A.K. Agrawal, Advancing Bridge Infrastructure Management through Artificial Intelligence: A Comprehensive Review, International Journal of Bridge Engineering, Management and Research, (2025).
[26] A. Krizhevsky, I. Sutskever, G.E. Hinton, ImageNet classification with deep convolutional neural networks, Communications of the ACM, 60 (2012) 84 - 90.
[27] Y.J. Cha, W. Choi, O. Büyüköztürk, Deep Learning-Based Crack Damage Detection Using Convolutional Neural Networks, Computer-Aided Civil and Infrastructure Engineering, 32 (2017).
[28] L. Chen, J. Zheng, Q. Chen, L. Jiang, T.D. Ngo, Semantic Segmentation Model for Road Cracks Based on Parallel Flatten Swin-VanillaNet Framework, Transportation Research Record, 2680 (2025) 793 - 813.
[29] J. Zheng, H. Lv, H. Song, J.-s. Li, R. Bai, L. Chen, Q. Chen, L. Jiang, FMANet: Fused mamba attention model with multi-type preprocessing for simulated crack-contaminated complex environments, Adv. Eng. Informatics, 69 (2026) 103808.
[30] Y. Ni, J. Mao, H. Wang, Z. Xi, Z. Chen, Surface Damage Detection and Localization for Bridge Visual Inspection Based on Deep Learning and 3D Reconstruction, Structural Control and Health Monitoring, (2024).
[31] G. Kim, Y. Cha, Deep learning-based 3D image reconstruction and damage mapping using neural radiance fields (Nerfacto), Structural Health Monitoring, (2025).
[32] X. Liang, Image-based post-disaster inspection of reinforced concrete bridge systems using deep learning with Bayesian optimization, Computer-Aided Civil and Infrastructure Engineering, 34 (2018) 415 - 430.
[33] R. Li, L. Zhao, H. Wei, G. Hu, Y. Xu, B. Ouyang, J. Tan, Multi-defect type beam bridge dataset: GYU-DET, Scientific Data, 12 (2025).
[34] J. Flotzinger, P.J. Rösch, N. Oswald, T. Braml, dacl1k: Real-World Bridge Damage Dataset Putting Open-Source Data to the Test, ArXiv, abs/2309.03763 (2023).
[35] Y. Zhang, K.V. Yuen, Review of artificial intelligence-based bridge damage detection, Advances in Mechanical Engineering, 14 (2022).
[36] H.S. Munawar, A.W.A. Hammad, S.T. Waller, M.R. Islam, Modern Crack Detection for Bridge Infrastructure Maintenance Using Machine Learning, Human-Centric Intelligent Systems, 2 (2022) 95 - 112.
[37] Y. Shi, L. Cui, Z. Qi, F. Meng, Z. Chen, Automatic Road Crack Detection Using Random Structured Forests, IEEE Transactions on Intelligent Transportation Systems, 17 (2016) 3434 - 3445.

[38] Q. Zou, Z. Zhang, Q. Li, X. Qi, Q. Wang, S. Wang, DeepCrack: Learning Hierarchical Convolutional Features for Crack Detection, IEEE Transactions on Image Processing, 28 (2019) 1498-1512.
[39] F. Yang, L. Zhang, S. Yu, D.V. Prokhorov, X. Mei, H. Ling, Feature Pyramid and Hierarchical Boosting Network for Pavement Crack Detection, IEEE Transactions on Intelligent Transportation Systems, 21 (2019) 1525-1535.
[40] V. Giglioni, J. Poole, I. Venanzi, F. Ubertini, K. Worden, On the use of domain adaptation techniques for bridge damage detection in a changing environment, ce/papers, 6 (2023).
[41] T.-Y. Lin, P. Dollár, R.B. Girshick, K. He, B. Hariharan, S.J. Belongie, Feature Pyramid Networks for Object Detection, 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), (2016) 936-944.
[42] J. Zhang, J. Li, R. Ly, Y. Liu, J.-H. Shu, Deep Learning-Based Fatigue Cracks Detection in Bridge Girders using Feature Pyramid Networks, ArXiv, abs/2410.21175 (2021).
[43] Y. Song, Q. Zhang, Y. Su, S. Zhang, R. Wang, W. Zhang, Z. Bi, Y. Yu, Advances in crack dataset development and deep learning-based detection models, Journal of Building Engineering, (2025).
[44] S. Dorafshan, R.J. Thomas, M. Maguire, SDNET2018: An annotated image dataset for non-contact concrete crack detection using deep convolutional neural networks, Data in brief, 21 (2018) 1664-1668.
[45] M. Mundt, S. Majumder, S. Murali, P. Panetsos, V. Ramesh, Meta-learning convolutional neural architectures for multi-target concrete defect classification with the concrete defect bridge image dataset, Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, 2019, pp. 11196-11205.
[46] J. Flotzinger, P.J. Rösch, T. Braml, dacl10k: benchmark for semantic bridge damage segmentation, Proceedings of the IEEE/CVF winter conference on applications of computer vision, 2024, pp. 8626-8635.
[47] R. Li, L. Zhao, H. Wei, G. Hu, Y. Xu, B. Ouyang, J. Tan, Multi-defect type beam bridge dataset: GYU-DET, Scientific Data, 12 (2025) 1101.
[48] E. Bianchi, A.L. Abbott, P. Tokekar, M. Hebdon, COCO-bridge: structural detail data set for bridge inspections, Journal of Computing in Civil Engineering, 35 (2021) 04021003.
[49] Z. Ameli, E. Landis, Corrosion condition rating database, (2023).
[50] Y. Tian, Q. Ye, D. Doermann, Yolov12: Attention-centric real-time object detectors, Advances in neural information processing systems, 38 (2026) 78433-78457.
[51] A. Chandrashekhar, B. Satyanarayana, R.R. Gorrepati, P. Vasanthi, K.L. Prasanna, An efficient YOLOv12-based framework for detecting extremely small-scale objects, Scientific Reports, (2025).
[52] X. Yin, C. Wang, W. Chen, Z. Zeng, Y. Quan, Z. Huang, A UAV-deployable lightweight framework for real-time bridge crack detection via YOLO-LY algorithm, Engineering Structures, 358 (2026) 122629.
[53] R. Bai, D. Luo, L. CHEN, X. Guo, H. Sun, Q. Chen, S. Qikai, L. Jiang, YOLO-CAB: a steel bridge cable damage detection model integrating multi-scale feature fusion and deformable convolution, Measurement Science and Technology, (2026).
[54] J. Dai, H. Qi, Y. Xiong, Y. Li, G. Zhang, H. Hu, Y. Wei, Deformable convolutional networks, Proceedings of the IEEE international conference on computer vision, 2017, pp. 764-773.
[55] S.J. Pan, Q. Yang, A survey on transfer learning, IEEE Transactions on knowledge and data engineering, 22 (2009) 1345-1359.
[56] H. Hu, J. Zhang, Y. Huang, R. Li, YOLO-GFE: Bolt detection and looseness angle estimation with geometry-guided feature enhancement, Displays, (2026) 103472.
[57] Y. Gu, D. Peng, J. Song, S. Ren, C. Kong, Image-based detection of bolts and bolt-missing defects

in multi-angle and complex background scenarios, Scientific Reports, (2026).
[58] L. Bao, S. Chen, Y. Bao, B. Li, J. Zhao, L. Yu, CrackLite: Lightweight Topology -Aware Crack Segmentation via Direction-Guided Topology Aggregation, (2026).
[59] J. Liu, W. Wang, H. Pu, Z. Cao, Y. Wang, H. Wang, K. Luo, Contour-Native Bridge Defect Detection and Compact Digital Archiving with Frequency-Supervised Fourier Contours, arXiv preprint arXiv:2605.08781, (2026).
[60] C. Lyu, S. Lin, A. Lynch, Y. Zou, M. Liarokapis, UAV-based deep learning applications for automated inspection of civil infrastructure, Automation in Construction, 177 (2025) 106285.