Image and Vision Computing 154 (2025) 105416

Contents lists available at ScienceDirect

Image and Vision Computing

journal homepage: www.elsevier.com/locate/imavis

AGSAM-Net: UAV route planning and visual guidance model for bridge surface defect detection

Rongji Li, Ziqian Wang [*]

Faculty of Information Technology, Monash University, Melbourne, 3800, Australia

ARTICLE INFO

Keywords:
Bridge surface defect detection
Unmanned aerial vehicle (UAV)
Multi-class semantic segmentation
Attention gating mechanism
AGSAM-net

ABSTRACT

Crack width is a critical indicator of bridge structural health. This paper proposes a UAV-based method for detecting bridge surface defects and quantifying crack width, aiming to improve efficiency and accuracy. The system integrates a UAV with a visual navigation system to capture high-resolution images (7322 × 5102 pixels) and GPS data, followed by image resolution computation and plane correction. For crack detection and segmentation, we introduce AGSAM-Net, a multi-class semantic segmentation network enhanced with attention gating to accurately identify and segment cracks at the pixel level. The system processes 8064 × 6048 pixel images in 2.4 s, with a detection time of 0.5 s per 540 × 540 pixel crack bounding box. By incorporating distance data, the system achieves over 90% accuracy in crack width quantification across multiple datasets. The study also explores potential collaboration with robotic arms, offering new insights into automated bridge maintenance.

1. Introduction

As global infrastructure rapidly expands, bridges, which serve as crucial nodes in transportation networks, play a pivotal role in ensuring the smooth operation of the economy and society [1,2]. However, as the service life of bridges extends and traffic loads increase, issues related to structural aging and damage become more pronounced. Common surface defects on bridges include honeycombing, seepage, repair marks, and spalling, with irregular cracks being particularly concerning as they can be early indicators of structural degradation [3]. If these defects are not promptly identified and addressed, they may lead to severe structural problems, potentially causing bridge collapse or other catastrophic failures. Consequently, the detection of surface defects and the health monitoring of bridges have become essential tasks in bridge maintenance. Traditional manual inspection methods are limited by their subjectivity, high repeatability, low efficiency, and significant safety risks [4]. Moreover, the accuracy of these inspections heavily depends on the experience and skill of the inspectors, making it challenging to meet the demands of modern bridge assessment [5].

To address these challenges, automated detection technologies based on computer vision and deep learning have gained significant attention in recent years, showing great promise for practical applications. Convolutional neural networks (CNNs) [6–8], a core technology in deep learning, have been widely employed in bridge surface defect detection. Classical semantic segmentation models such as U-Net [9], SegNet [10], and the Deeplabv3+ [11] series have been used to accurately classify and segment defects like cracks, spalling, and corrosion at the pixel level on bridge surfaces. U-Net, with its unique encoder–decoder architecture and skip connections, provides efficient segmentation suitable for complex bridge inspection tasks. SegNet enhances segmentation performance and reduces model complexity through its pooling index mechanism, while the Deeplabv3+ series leverages dilated convolutions and the atrous spatial pyramid pooling (ASPP) module to significantly expand the receptive field and adapt to multi-scale targets [12,13]. Despite their strong laboratory performance, these models face several challenges in real-world applications, such as poor detection of small defects, susceptibility to complex background interference, and inaccurate defect classification.

To further improve the precision and robustness of bridge defect detection, researchers have proposed various enhanced methods. For instance, deep learning models integrating attention mechanisms have shown significant improvements in detection accuracy by focusing on key areas and filtering out redundant information. Additionally, non-contact detection technologies have been introduced to the field of bridge inspection to overcome the limitations of traditional image processing methods in complex scenarios. These technologies, including infrared thermography, ground-penetrating radar (GPR) [14], and terrestrial laser scanning (TLS) [15], enable the rapid acquisition of high-resolution surface images and internal structural data without disrupting traffic. Such multi-source data provide comprehensive support for bridge health monitoring [16]. However, effectively integrating

[*] Corresponding author.
E-mail addresses: richard998@yeah.net (R. Li), snowblindabc@163.com (Z. Wang).

https://doi.org/10.1016/j.imavis.2025.105416
Received 27 August 2024; Received in revised form 14 November 2024; Accepted 7 January 2025
Available online 13 January 2025

these diverse data sources with advanced machine learning algorithms to achieve efficient, accurate, and real-time bridge defect detection remains a key research focus.

Against this backdrop, this study proposes an unmanned aerial vehicle (UAV)-based bridge surface defect detection system [17,18], which integrates visual guidance, path planning, and multi-class semantic segmentation capabilities [19]. The system utilizes an improved multi-class semantic segmentation network, AGSAM-Net, combined with an attention gating mechanism and the ASPP module to enhance the detection capabilities for bridge defects. By leveraging the attention mechanism, the system can accurately identify and locate various bridge surface defects in complex backgrounds, achieving precise segmentation of small cracks. Additionally, the system incorporates automatic path planning functionality for the UAV, enabling efficient inspection of large bridge areas with minimal human intervention. Experimental results demonstrate that the system exhibits excellent detection accuracy and generalization capabilities across multiple bridge datasets, particularly excelling in the accuracy and real-time detection of bridge surface defects.

This research aims to develop a UAV-based bridge surface defect detection system that integrates visual guidance and route planning to achieve efficient and accurate bridge surface inspections. The system will employ a multi-class semantic segmentation network combined with attention mechanisms to significantly enhance the detection of small defects in complex backgrounds. Beyond automated detection, the study explores the potential integration of this system with an embodied robotic arm to form a comprehensive solution for automated bridge defect detection, evaluation, and welding repair, providing holistic support for bridge maintenance operations.

Our main contributions are as follows:

1. By integrating attention gating mechanisms and the ASPP module, the system significantly improves the accuracy and robustness of detecting bridge surface defects, particularly in recognizing small cracks and defects within complex backgrounds.
2. The combination of path planning optimization and visual guidance technology enables the system to efficiently cover large bridge areas and precisely locate defects, reducing human intervention and enhancing detection efficiency.
3. A conceptual model for collaborative detection and welding repair using a visually guided robotic arm is proposed, laying the groundwork for future automated bridge repair technologies. This research takes a crucial step toward realizing intelligent and automated bridge maintenance.

2. Related work

2.1. Application of UAVs in bridge defect detection

In recent years, unmanned aerial vehicles (UAVs) have gained significant traction in bridge inspection due to their flexibility, efficiency, and capability to access hard-to-reach areas [20]. UAVs eliminate the need for scaffolding or lane closures, allowing for rapid inspection of various parts of a bridge while minimizing disruption to traffic [21]. This non-contact method also enhances safety by reducing the need for human inspectors to operate in hazardous environments, such as those found in high or structurally compromised bridges [22]. However, despite their advantages, existing UAV inspection systems face several limitations in path planning and data acquisition, which hinder their broader application in bridge inspection.

One of the primary challenges in UAV-based bridge inspection is achieving efficient coverage of complex bridge structures [23]. Traditional path planning algorithms often struggle to navigate the intricate geometries of bridges, leading to incomplete inspections or the need for multiple passes, which reduces overall efficiency. Moreover, these algorithms may not consider the unique characteristics of different bridge types, such as suspension bridges, arch bridges, or cable-stayed bridges, each of which requires tailored inspection strategies.

Another significant limitation is the lack of precise positioning information during data collection. Many UAV systems rely on standard GPS technology, which may not provide the accuracy needed for detailed structural assessments [23]. Without high-precision GPS or alternative localization methods, the collected images may suffer from misalignment, making it difficult to accurately map defects to specific locations on the bridge. This misalignment can lead to challenges in post-processing, such as stitching images together or identifying areas that require further inspection.

To address these issues, recent advancements have focused on improving path planning algorithms and integrating high-precision GPS technologies. For instance, enhanced path planning algorithms that incorporate 3D models of bridges have been developed to ensure more comprehensive coverage. These algorithms consider the bridge's unique structure and optimize the UAV's flight path to minimize blind spots and reduce inspection time. Additionally, the integration of Real-Time Kinematic (RTK) GPS and other advanced localization technologies has significantly improved the accuracy of positioning, enabling precise defect mapping and facilitating more reliable data collection [6].

The combination of improved path planning and precise GPS localization not only enhances the UAV's coverage and data acquisition capabilities but also provides a more robust foundation for subsequent defect detection and analysis. These advancements make UAVs a more viable solution for large-scale bridge inspections, particularly in complex or hazardous environments where traditional methods are impractical or too risky.

2.2. Development of automated defect detection technologies

Traditional bridge defect detection has long relied on manual visual inspection, a process that is inherently subjective, labor-intensive, and prone to human error [24]. The limitations of manual inspections include inconsistent results due to varying inspector expertise, the inability to detect small or subtle defects, and the slow pace of inspections, which can be a significant drawback for large-scale infrastructure. These challenges have driven the development of automated defect detection technologies, particularly those based on deep learning and computer vision [25].

CNNs have emerged as a powerful tool for automated defect detection [26]. CNNs excel at feature extraction and classification tasks, making them ideal for identifying various types of bridge surface defects, such as cracks, spalling, and corrosion. Semantic segmentation, a technique that classifies each pixel in an image into a category, has been widely adopted in bridge inspection to achieve detailed and accurate defect mapping [27]. Models such as U-Net, SegNet, and Deeplabv3+ have become the foundation of many automated inspection systems.

U-Net [28], with its encoder–decoder architecture and skip connections, has been particularly effective in semantic segmentation tasks. The model's ability to capture both high-level context and fine-grained details makes it well-suited for identifying defects in complex environments. SegNet, on the other hand, reduces computational complexity by using pooling indices to upsample features, making it more efficient for real-time applications. The Deeplabv3+ series introduces dilated convolutions and the ASPP module, which expands the receptive field without losing resolution, thereby improving the model's ability to detect defects of varying sizes.

Despite these advancements, existing models still face challenges when dealing with complex backgrounds and small defects [29]. For instance, in scenarios where the bridge surface has significant texture or varying lighting conditions, CNNs may misclassify defects or miss them altogether. Additionally, small or irregular defects, which are often the most critical indicators of structural issues, are difficult for these models to detect consistently. To address these challenges, recent research has explored the integration of attention mechanisms, which allow models to focus on relevant areas of an image, and multi-scale feature fusion techniques, which improve the detection of defects across different scales.

The ongoing development of automated defect detection technologies is crucial for enhancing the accuracy, reliability, and efficiency of bridge inspections. By overcoming the limitations of traditional methods, these technologies offer the potential for continuous monitoring and real-time assessment of bridge conditions, ultimately contributing to the safety and longevity of critical infrastructure.

2.3. Application of visual guidance technology in bridge inspection

Visual guidance technology plays a critical role in enabling automated inspection systems by providing real-time analysis and interpretation of visual data [30]. In bridge inspection, visual guidance systems are used to direct UAVs and other inspection robots, ensuring they capture relevant data and identify potential defects accurately [31]. These systems rely on advanced image processing algorithms to analyze the visual information collected during inspections, guiding the UAV or robot along optimal paths and highlighting areas that require closer examination.

However, existing visual guidance systems face challenges in terms of processing speed and accuracy, especially when dealing with complex environments and diverse types of defects. Bridges often have varying surface textures, lighting conditions, and background interference, which can complicate defect detection [32]. Furthermore, the presence of different types of defects, such as cracks, spalling, or corrosion, requires the system to accurately classify and prioritize these issues, which can be difficult with current technology.

To enhance the performance of visual guidance systems, recent advancements have focused on improving image preprocessing workflows and algorithm architectures. For example, adjustments in brightness, contrast, and noise reduction can significantly improve the quality of the input images, making it easier for the system to detect and classify defects [33]. Additionally, image segmentation techniques that divide the image into smaller, more manageable sections have been shown to improve the accuracy of defect detection, particularly in cases where defects are small or irregularly shaped.

Moreover, the integration of machine learning algorithms with visual guidance systems has opened new possibilities for autonomous bridge inspection [34]. By training models on large datasets of bridge images, these systems can learn to recognize a wide range of defect types and adjust their inspection strategies accordingly. For instance, a system might learn to prioritize areas of a bridge that are more prone to cracking based on historical data, thereby optimizing the inspection process.

The application of visual guidance technology extends beyond inspection; it also has significant potential in automated repair systems. By integrating visual guidance with robotic arms, it is possible to develop systems that not only detect but also repair defects autonomously. For example, a robotic arm equipped with a welding tool could be guided by visual data to precisely repair cracks or reinforce weakened areas [35]. This integration could lead to the development of comprehensive systems that handle both detection and repair, greatly reducing the need for human intervention and improving the efficiency of bridge maintenance operations.

In summary, visual guidance technology is a cornerstone of modern bridge inspection systems, enabling more accurate and efficient defect detection. By addressing the current challenges of speed and accuracy through advancements in image processing and machine learning, these systems are becoming increasingly capable of handling the complexities of real-world bridge inspections. The future integration of visual guidance with automated repair technologies promises to further revolutionize the field, offering a holistic solution for maintaining and extending the lifespan of critical infrastructure.

3. Our approach

AGSAM-Net is a multi-class semantic segmentation network designed specifically for detecting surface defects on bridges. The network architecture of AGSAM-Net is shown in Fig. 1. The network integrates an attention gating mechanism and an atrous spatial pyramid pooling (ASPP) module [36] to enhance its accuracy in detecting cracks and other defects in complex environments. AGSAM-Net employs a classic encoder–decoder architecture and uses skip connections to achieve multi-scale feature fusion, enabling the effective capture of both fine details and contextual information. The primary advantage of AGSAM-Net lies in its attention gating mechanism, which allows the network to focus more on critical areas such as cracks while suppressing background noise. Additionally, the use of atrous convolutions and the ASPP module further expands the receptive field, thereby improving the network's generalization capability and robustness.

Before the detection process begins, the bridge surface images captured by the drone undergo pre-processing, including resolution correction, contrast adjustment, and noise removal. These steps are essential to ensure that the subsequent feature extraction and segmentation processes receive clear and precise input data. The encoder part of AGSAM-Net consists of multiple convolutional layers, each followed by batch normalization and a ReLU activation function. Max pooling is applied to progressively reduce the spatial dimensions of the image, extracting features at different scales. During this process, the encoder gradually captures high-level features of cracks and other defects in the image. At the end of the encoder, AGSAM-Net incorporates an ASPP module, which captures multi-scale information through convolutions with various dilation rates. This significantly expands the network's receptive field, allowing it to retain more contextual information without increasing computational complexity.

AGSAM-Net integrates an attention gating mechanism to enhance the focus on critical regions. In the decoder, the attention module automatically adjusts the focus on each region based on the information passed from the encoder. By calculating the correlation between feature maps and a query vector, the network can automatically filter out irrelevant background information and allocate resources to process the key crack regions. The decoder gradually restores the resolution of the feature maps to match the original image, using upsampling and convolutional operations. Through skip connections, high-resolution features from the encoder are combined with those in the decoder, ensuring that fine details are preserved during the upsampling process. In the final layer of the decoder, AGSAM-Net uses a Softmax function to classify each pixel, outputting a probability distribution for different defect categories. The resulting segmentation map includes classifications for cracks, seepage, spalling, and other bridge surface defects.

Finally, the system combines the drone's GPS data with image resolution information to quantify the width and length of detected cracks. These quantifications are used for further analysis of the bridge's health and maintenance planning. Through these steps, AGSAM-Net is capable of achieving precise multi-class semantic segmentation in complex bridge scenarios, providing an efficient and robust solution for the automated detection of bridge surface defects.

3.1. AGSAM-net model structure

AGSAM-Net is a multi-class semantic segmentation network designed for detecting surface defects on bridges. It enhances the classic U-Net architecture to improve the capture of multi-scale features and detail processing in complex scenarios. The basic architecture of AGSAM-Net follows the encoder–decoder structure of U-Net, where the encoder gradually extracts image features, and the decoder progressively restores image resolution. The network utilizes skip connections to achieve multi-scale feature fusion. Additionally, AGSAM-Net incorporates Dilated Convolution and an Atrous Spatial Pyramid Pooling

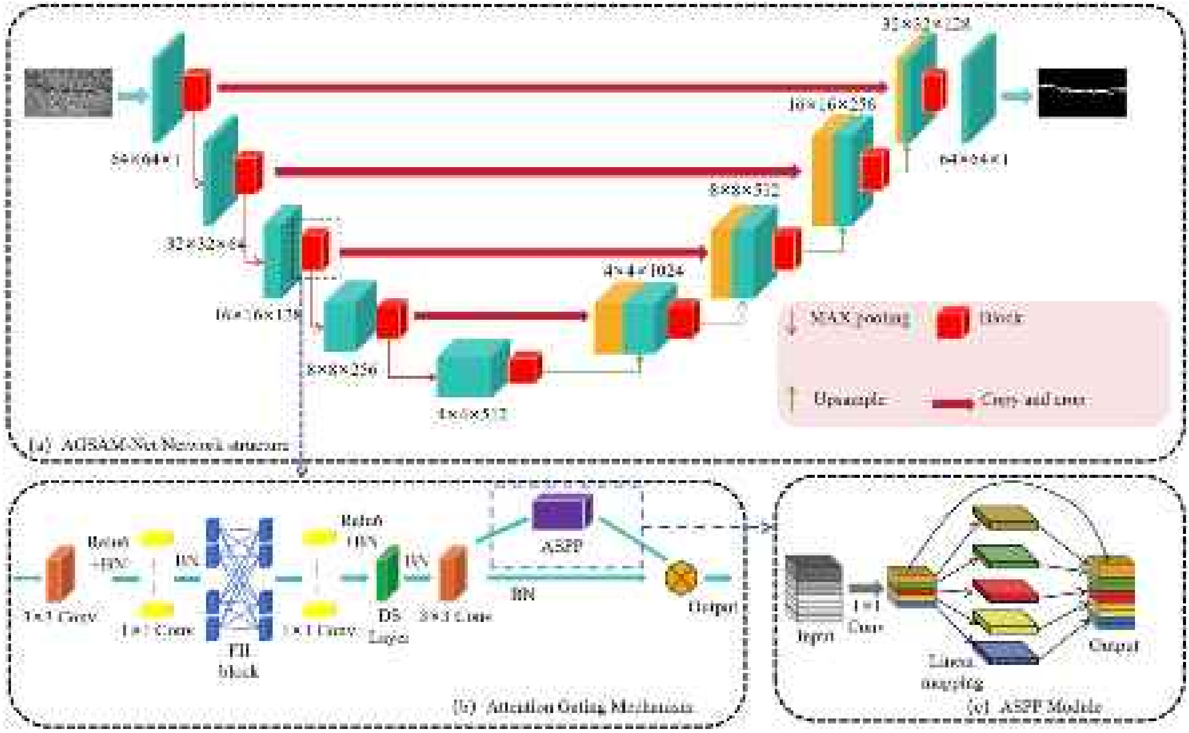

Fig. 1. Our network structure. (a) The network structure of AGSAM-Net; (b) Attention Gating Mechanism model structure; (c) ASPP Module.

(ASPP) module to expand the network's receptive field and improve its sensitivity to multi-scale information.

In the encoder part, the image is processed layer by layer through a series of Convolution (Conv) operations, Batch Normalization (BN), and ReLU activation functions. Each encoding layer consists of two convolutional layers with 3 × 3 kernels, followed by a 2 × 2 Max Pooling layer for downsampling. The output feature maps of each layer are reduced in resolution via max pooling, extracting higher-level features. The computation for each convolutional layer in the encoder is as follows:

$$F_{enc}^{l} = \mathrm{ReLU}(\mathrm{BN}(\mathrm{Conv}(F_{enc}^{l-1}))), \tag{1}$$

where F_{enc}^{l} represents the output feature map of the lth layer of the encoder, Conv denotes the 3 × 3 convolution operation, BN denotes batch normalization, and ReLU is the activation function.

In the decoder part, the feature maps are gradually upsampled to restore the original image resolution. The decoder and encoder use skip connections, where high-resolution features from the encoder are directly passed to the corresponding layer in the decoder for feature fusion. The computation for each decoder layer is as follows:

$$F_{dec}^{l} = \mathrm{ReLU}(\mathrm{BN}(\mathrm{Conv}(\mathrm{Concat}(F_{dec}^{l+1}, F_{enc}^{l})))), \tag{2}$$

where F_{dec}^{l} represents the output feature map of the lth layer of the decoder, and Concat represents the concatenation operation.

To extend the receptive field, AGSAM-Net introduces Dilated Convolution and an ASPP module between the encoder and decoder. Dilated Convolution uses a dilation rate to enlarge the receptive field of the convolutional kernel, capturing more contextual information without increasing the number of parameters. The computation for Dilated Convolution is given by:

$$F_{dilated} = \sum_{i=1}^{k} w_i \cdot x_{r(i)}, \tag{3}$$

where w_i represents the weights of the convolutional kernel, $x_{r(i)}$ denotes the input signal, and $r(i)$ denotes the index with dilation rate.

The ASPP module captures multi-scale information through convolutions with different dilation rates and concatenates these outputs to enhance the network's ability to detect defects of various sizes and shapes. The output of the ASPP module is expressed as:

$$F_{ASPP} = \mathrm{Concat}(\mathrm{Conv}_1, \mathrm{Conv}_2, \ldots, \mathrm{Conv}_n), \tag{4}$$

where Conv_i represents convolutions with different dilation rates, and n is the number of dilation rates.

AGSAM-Net's skip connections concatenate the high-resolution feature maps from the encoder with the corresponding feature maps in the decoder, achieving multi-scale feature fusion. Through these connections, the decoder can retain both high contextual information and edge detail, thereby improving segmentation accuracy. The fusion operation for the skip connection is described by:

$$F_{fusion} = \mathrm{Concat}(F_{encoder}, F_{decoder}), \tag{5}$$

where $F_{encoder}$ denotes the high-resolution feature map from the encoder, and $F_{decoder}$ denotes the feature map from the current decoder layer.

Finally, after a series of upsampling and convolution operations, AGSAM-Net uses the Softmax function to classify each pixel, outputting a multi-class probability distribution. The Softmax function is defined as:

$$P(c|x) = \frac{\exp(f_c(x))}{\sum_{j=1}^{C} \exp(f_j(x))}, \tag{6}$$

where $P(c|x)$ is the probability of the pixel belonging to class c, $f_c(x)$ is the score for class c, and C is the total number of classes. With these improvements, AGSAM-Net effectively handles multi-scale and complex background issues in bridge surface defect detection, significantly enhancing segmentation accuracy and generalization capability.

3.2. Attention gating mechanism

The Attention Gating Mechanism plays a critical role in deep learning, particularly in the field of computer vision. Its basic principle is similar to the human attention mechanism, which dynamically allocates computational resources to focus on the most important parts of

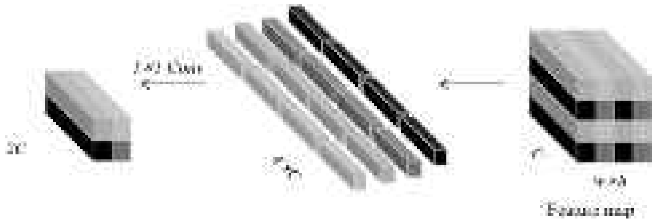

Fig. 2. Schematic diagram of the DS layer.

the input data. In semantic segmentation tasks, the attention mechanism selectively enhances the response to target regions while suppressing background information, thereby improving the network's ability to capture details and segmentation accuracy.

In multi-class semantic segmentation tasks, the importance of different pixels is typically uneven. Traditional CNNs treat all input information equally, leading to inefficient use of computational resources and reduced model performance. To address this, the Attention Gating Mechanism calculates the correlation between input feature maps and contextual information, automatically assigning attention weights, allowing the network to focus more on target regions while ignoring irrelevant areas. The introduction of this mechanism enables the model to dynamically adjust its focus on each pixel, thereby enhancing segmentation accuracy and robustness.

The basic idea of the Attention Gating Mechanism is to calculate the importance of each position in the feature map (i.e., attention weights) and reweight the feature map according to these weights to highlight key regions. During the computation, the input feature map is first linearly transformed to extract attention-related features. Then, the similarity between the feature map and a given query vector is calculated to obtain attention weights. Finally, these weights are used to reweight the original feature map to generate a context vector.

Assume the input feature map is $\mathbf{X} = [\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n]$, where $\mathbf{x}_i$ represents the feature vector of the ith pixel, and the query vector is $\mathbf{q}$. The attention score for the ith pixel can be expressed as:

$$e_i = f(\mathbf{x}_i, \mathbf{q}), \tag{7}$$

where $f(\mathbf{x}_i, \mathbf{q})$ is the function used to compute similarity, typically using a dot product or additive function. Next, the attention weights for each pixel are obtained by normalizing all scores using the Softmax function:

$$\alpha_i = \frac{\exp(e_i)}{\sum_{j=1}^{n} \exp(e_j)}, \tag{8}$$

where α_i represents the attention weight of the ith pixel. The final context vector $\mathbf{c}$ is obtained by the weighted sum of all input feature maps:

$$\mathbf{c} = \sum_{i=1}^{n} \alpha_i \mathbf{x}_i, \tag{9}$$

where $\mathbf{c}$ represents the feature vector of the attended region.

The DS layer functions similarly to a downsampling operation, receiving output features from the last layer of the previous convolutional operation. As shown in Fig. 2, it concatenates each adjacent 2 × 2 block along the channel direction within the feature map. A subsequent 1 × 1 convolution then refines the channel dimension of the concatenated feature map. This process effectively reduces the resolution of the feature map by half while simultaneously doubling the channel size from C to $2C$. In AGSAM-Net, the DS layer's operation is integrated into the architecture to efficiently handle multi-scale features. The concatenation of adjacent blocks and the dimensional refinement enhance the network's ability to capture finer details at reduced resolutions, thereby contributing to improved segmentation accuracy in identifying bridge surface defects such as cracks. The refined and reduced feature maps help the model focus on critical regions while suppressing background noise, ensuring high precision in complex scenarios.

Specifically, AGSAM-Net first extracts multi-level feature maps through the encoder and then applies the Attention Gating Mechanism in the decoder stage to process these feature maps. For each layer of the feature map, the model calculates its correlation with the target region and reweights the feature map according to the calculated attention weights. This process gradually enhances the response to key regions such as cracks during decoding, thereby improving the model's segmentation accuracy.

The Attention Gating Mechanism not only allows AGSAM-Net to more accurately identify and segment bridge surface defects in complex scenarios but also effectively suppresses background noise, enhancing the model's robustness. Thus, the introduction of the attention mechanism is a key factor in achieving high precision and stable performance in AGSAM-Net's practical applications.

3.3. Loss function design

Algorithm 1: Softmax Loss Calculation in AGSAM-Net

```
Input: Feature map $\mathbf{X} \in \mathbb{R}^{H \times W \times C}$, true labels $\mathbf{Y} \in \{1, \ldots, K\}^{H \times W}$
Output: Softmax Loss $\mathcal{L}$
for each pixel $i \in \mathbf{X}$ do
    for each class $j \in \{1, \ldots, K\}$ do
        Compute logits: $z_{i,j} = \mathbf{W}_j^{\top} \mathbf{x}_i + b_j$;
        Compute softmax probability:
        $p_{i,j} = \frac{\exp(z_{i,j})}{\sum_{k=1}^{K} \exp(z_{i,k})}$    (10)
    end
    Compute cross-entropy loss for pixel $i$:
    $\ell_i = -\sum_{j=1}^{K} y_{i,j} \log(p_{i,j})$    (11)
end
Compute average Softmax Loss across all pixels:
$\mathcal{L} = \frac{1}{N} \sum_{i=1}^{N} \ell_i$    (12)
return $\mathcal{L}$
```

In multi-class semantic segmentation tasks, the design of the loss function is critical to the model's performance. The primary role of the loss function is to measure the difference between the predicted results and the ground truth labels, guiding the model parameters' update through backpropagation to improve accuracy. In the context of bridge surface defect detection, the sample distribution across different defect types is often imbalanced. Traditional loss functions may cause the model to bias toward predicting the majority class. To address the class imbalance and improve segmentation accuracy for minority classes, AGSAM-Net employs a combination of Cross-Entropy Loss [37] and Softmax Loss [38].

Cross-Entropy Loss is commonly used in classification tasks, aiming to minimize the divergence between the true distribution and the predicted distribution. For multi-class problems, Cross-Entropy Loss effectively measures the deviation between the model's predicted probability distribution and the ground truth. However, in scenarios with imbalanced data, the contribution of the minority classes to the overall loss is minimal, which may lead to inaccurate predictions for those classes.

To mitigate this issue, we incorporate Softmax Loss. Softmax Loss penalizes the misclassification of hard-to-classify samples more heavily, thereby increasing the model's focus on these samples and improving classification accuracy. Specifically, Softmax Loss strengthens the gradient of minority class samples during training, assigning higher weights to these samples to alleviate the impact of class imbalance on model performance.

In a multi-class problem, given a set of samples $\mathbf{x} = \{x_1, x_2, \ldots, x_n\}$ with corresponding true labels $\mathbf{y} = \{y_1, y_2, \ldots, y_n\}$, the class probabilities are computed using the Softmax function, expressed as:

$$p_{i,j} = \frac{\exp(z_{i,j})}{\sum_{k=1}^{K} \exp(z_{i,k})}, \tag{13}$$

where $p_{i,j}$ represents the probability that the ith sample is classified as the jth class, $z_{i,j}$ is the model's output score for class j, and K is the total number of classes.

The loss for the multi-class problem is calculated as:

$$\text{Loss} = -\frac{1}{n} \sum_{i=1}^{n} \sum_{j=1}^{K} y_{i,j} \log(p_{i,j}), \tag{14}$$

where $y_{i,j}$ is the true label for the ith sample, with $y_{i,j} = 1$ if the sample belongs to class j, otherwise 0. Softmax Loss extends Cross-Entropy Loss by incorporating a penalty for hard-to-classify samples. Let p_c represent the predicted probability for class $c \in C$, calculated as:

$$p_c = \frac{\exp(z_c)}{\sum_{k=1}^{K} \exp(z_k)}, \tag{15}$$

where z_c is the score for the current class. When class c is a minority or challenging to distinguish, Softmax Loss automatically applies a weight to it, ensuring that the final loss gives greater importance to minority class samples. The Softmax Loss is defined as:

$$\text{Loss} = -\frac{1}{n} \sum_{i=1}^{n} \sum_{j=1}^{K} w_j y_{i,j} \log(p_{i,j}), \tag{16}$$

where w_j is the weight for class j, adjusting the contribution of different classes to the overall loss.

In AGSAM-Net, the loss function design, which combines Cross-Entropy Loss and Softmax Loss, effectively addresses the class imbalance issue in bridge surface defect detection. Cross-Entropy Loss ensures the model's overall classification capability, while Softmax Loss penalizes the misclassification of minority classes, allowing the model to more accurately identify defects such as cracks and spalling. This loss function design enables AGSAM-Net to maintain high accuracy and robustness in complex scenarios, especially in cases of small samples and imbalanced data.

Algorithm 1 describes the Softmax loss computation process used for multi-class semantic segmentation in AGSAM-Net. The input consists of feature maps $\mathbf{X}$ and corresponding ground truth labels $\mathbf{Y}$, and the output is the Softmax loss value $\mathcal{L}$. For each pixel, the logarithmic odds $z_{i,j}$ for each class are computed, followed by the probability $p_{i,j}$ through the Softmax function, and the pixel-wise loss ℓ_i is calculated using the cross-entropy formula. Finally, the Softmax loss value $\mathcal{L}$ is the mean of all pixel losses.

4. Experiment

4.1. Datasets

The Crack500 dataset [39] is a publicly available collection aimed at bridge crack detection, consisting of 500 labeled crack images, each with a resolution of approximately 2000 × 1500 pixels. To increase the dataset size and diversify the data, each image was cropped into 16 non-overlapping regions, retaining only those segments containing more than 1000 crack pixels. This preprocessing step expanded the dataset to include training, validation, and test sets, making it a valuable resource for evaluating crack detection models under varied conditions.

To further enhance data diversity, data augmentation techniques were applied, such as flipping, rotation, scaling, saturation adjustment, contrast variation, brightness control, and grayscale conversion. Fig. 3 illustrates the application of data augmentation to the bridge fracture dataset. These transformations increased the dataset size to approximately 4000 images, each resized to 540 × 540 pixels for defect recognition model training and validation. The dataset was split into training (70%), validation (20%), and testing (10%) subsets, with 2800 images used for training, 800 for validation, and 400 for testing. Model performance on the validation set was evaluated after each epoch using the Intersection over Union (IoU) metric to assess defect segmentation accuracy.

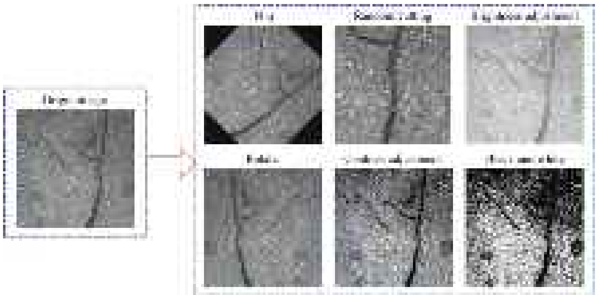

Fig. 3. Enhanced display of bridge pavement crack image data.

Fig. 4. The process of bridge crack image taken by UAV is shown.

Table 1
UAV system specifications.

Equipment	Specifications
UAV	DJI Mini 3
Maximum photo size	8064 × 6048 pixels
Lens	24 mm
Takeoff weight	248 g
Maximum flight time	51 min
Pan-tilt range	−135° to 80°
Pan-tilt roll range	−135° to 45°
Pan-tilt tilt range	−30° to 30°

The UAV-based dataset was employed for noise barrier data collection, serving as a foundation for training and validation. A hybrid feature learning approach was proposed for crack identification and width quantification using machine vision. To validate the proposed method, experimental studies on UAV-based bridge crack detection were conducted, and results were compared with established networks. Final manual measurements using a bridge inspection vehicle confirmed the accuracy of the method. UAV system specifications are detailed in Table 1. The initial collection consisted of approximately 300 high-resolution images of bridge pavement cracks that varied in width and posed a significant challenge for accurate identification, part of which is shown in Fig. 4. Table 2 provides details of the bridge pavement defect dataset.

4.2. Parameter configuration

The experimental setup is detailed in Table 3. The experiments were conducted on a system running Ubuntu 20.04, equipped with a 13th

Table 2
Bridge pavement defect dataset information.

Items	Crack500 (After augmentation)	UAV generalization
Number of images	4000	2400
Train	2800	1600
Test	800	480
Validation	400	240

Table 3
Experimental software and hardware configuration.

Configuration information	Version information
GPU	NVIDIA GeForce RTX 4090
Video memory	64G
CPU	13th Gen Intel(R) Core(TM) i7-13700KF
Internal memory	256G
Operating system	Ubuntu 20.04
Programing language	Python 3.8
Network framework	Pytorch 1.11.0
CUDA	11.3

Table 4
Experimental parameter configuration.

Parameter	Configuration
Epochs	100
Batch Size	16
Optimizer	Adam
Learning Rate	0.001
Weight Decay	1.00×10^{-4}
Loss Function	Softmax Loss

Gen Intel(R) Core(TM) i7-13700KF processor and 64 GB of memory. The deep learning model was implemented using Python 3.8, Pytorch 1.11.0, and CUDA 11.3. The training was performed on an NVIDIA GeForce RTX 4090 GPU with 64 GB of video memory. The model was trained for 100 epochs with a batch size of 16. The learning rate was initially set to 0.001 and adjusted using a cosine annealing schedule, which gradually reduced the learning rate following a cosine function over the course of the training. The optimizer used was Adam, with a weight decay set to 1×10^{-4}. The Focal Loss function was employed as the criterion for training. The experimental parameter configuration information is shown in Table 4.

4.3. Evaluating indicator

In semantic segmentation tasks, evaluation metrics play a crucial role, especially in the context of bridge defect detection. These metrics are essential for accurately quantifying the segmentation performance of models, which helps assess their reliability and effectiveness in real-world applications. For instance, Pixel Accuracy (PA) and Class Pixel Accuracy (CPA) evaluate the model's performance at both the overall pixel level and for specific classes. Meanwhile, Mean Class Pixel Accuracy (MPA) assesses the model's balanced performance across all classes. Intersection over Union (IoU) is another critical metric used to measure the overlap between the predicted results and the ground truth annotations. Additionally, Mean Intersection over Union (MIoU) and Frequency-Weighted Intersection over Union (FWIoU) extend the applicability of IoU by evaluating the model's performance in multi-class segmentation tasks. Recall and F1-Score further provide insights into the model's sensitivity to positive samples and the balance between precision and recall, both of which are vital in defect detection, where high sensitivity is often required.

Specifically, Pixel Accuracy is calculated as shown in Eq. (17), by measuring the ratio of correctly classified pixels to the total number of pixels. Class Pixel Accuracy, shown in Eq. (18), evaluates the accuracy for each class individually. The Mean Class Pixel Accuracy, given by Eq. (19), calculates the average accuracy across all classes. The Intersection over Union (Eq. (20)) measures the ratio of the intersection area to the union area between the predicted and actual segments. The Mean Intersection over Union (Eq. (21)) and Frequency-Weighted Intersection over Union (Eq. (22)) provide averaged and weighted IoU metrics, respectively, to assess overall segmentation quality. Recall (Eq. (23)) and F1-Score (Eq. (24)) are particularly important in detecting defects, ensuring that the model accurately identifies potential issues.

Accurate computation of these evaluation metrics is vital for ensuring the robustness and reliability of bridge defect detection systems. These metrics not only help the model perform well in controlled environments but also ensure its stability and effectiveness in practical bridge inspection scenarios. By utilizing these metrics, researchers can comprehensively evaluate the adaptability and robustness of models across different scenarios, ultimately enhancing the automation and safety management of bridge maintenance.

$$PA = \frac{\sum_{i=1}^{n} TP_i}{\sum_{i=1}^{n} (TP_i + FP_i)} \tag{17}$$

$$CPA_i = \frac{TP_i}{TP_i + FP_i} \tag{18}$$

$$MPA = \frac{1}{n} \sum_{i=1}^{n} CPA_i \tag{19}$$

$$IoU_i = \frac{TP_i}{TP_i + FP_i + FN_i} \tag{20}$$

$$MIoU = \frac{1}{n} \sum_{i=1}^{n} IoU_i \tag{21}$$

$$FWIoU = \frac{\sum_{i=1}^{n} |G_i| \cdot IoU_i}{\sum_{i=1}^{n} |G_i|} \tag{22}$$

$$Recall_i = \frac{TP_i}{TP_i + FN_i} \tag{23}$$

$$F1_i = \frac{2 \cdot Precision_i \cdot Recall_i}{Precision_i + Recall_i} \tag{24}$$

Through these well-designed evaluation metrics, we can comprehensively assess the performance of bridge defect detection models, thereby ensuring their reliability and efficiency in practical applications.

4.4. Contrast experiment

In this comparison experiment, we aim to verify the superiority of AGSAM-Net in semantic segmentation tasks. We selected the Crack500 dataset, which contains road surface crack segmentation images processed through data augmentation, as our testing dataset. We evaluated the performance of several mainstream semantic segmentation models, including U-Net, SegNet, FCN, Deeplabv3+, and PSPNet. The evaluation metrics primarily include Recall (%), F1-Score (%), MIoU (%), FWIoU (%), PA (%), and MPA (%). The experimental results are shown in Table 5.

The experimental results indicate that AGSAM-Net outperforms other models on the Crack500 dataset in terms of overall performance. Specifically, AGSAM-Net achieves the highest MIoU (%) of 75.4% and FWIoU (%) of 95.8%, surpassing the performance of the other models. Additionally, AGSAM-Net shows significant advantages in MPA, Recall, and F1-Score, while only being 0.2% behind PSPNet in PA. Compared to the second-best model, Deeplabv3+, AGSAM-Net exhibits improvements across all metrics, demonstrating its superior generalization capabilities in complex scenarios.

We also show the results of different algorithms, and the test results are shown in Fig. 5. It can be seen from the figure that our model achieves the best segmentation effect in crack detection, perfectly matching the crack width.

Table 5
Performance comparison of different semantic segmentation networks on the Crack500 dataset.

Method	Recall (%)	F1-Score (%)	MIoU (%)	FWIoU (%)	PA (%)	MPA (%)
U-Net [9]	82.1	82.4	70.4	95.2	96.5	79.5
SegNet [40]	82.4	81.2	71.2	95.5	96.3	81.2
FCN [41]	80.6	80.6	73.8	94.6	96.1	81.1
Deeplabv3+ [42]	81.5	82.3	74.5	95.4	95.7	80.6
PSPNet [43]	86.6	82.3	74.1	95.1	97.1	82.3
AGSAM-Net	86.4	85.2	75.4	95.8	97.4	85.4

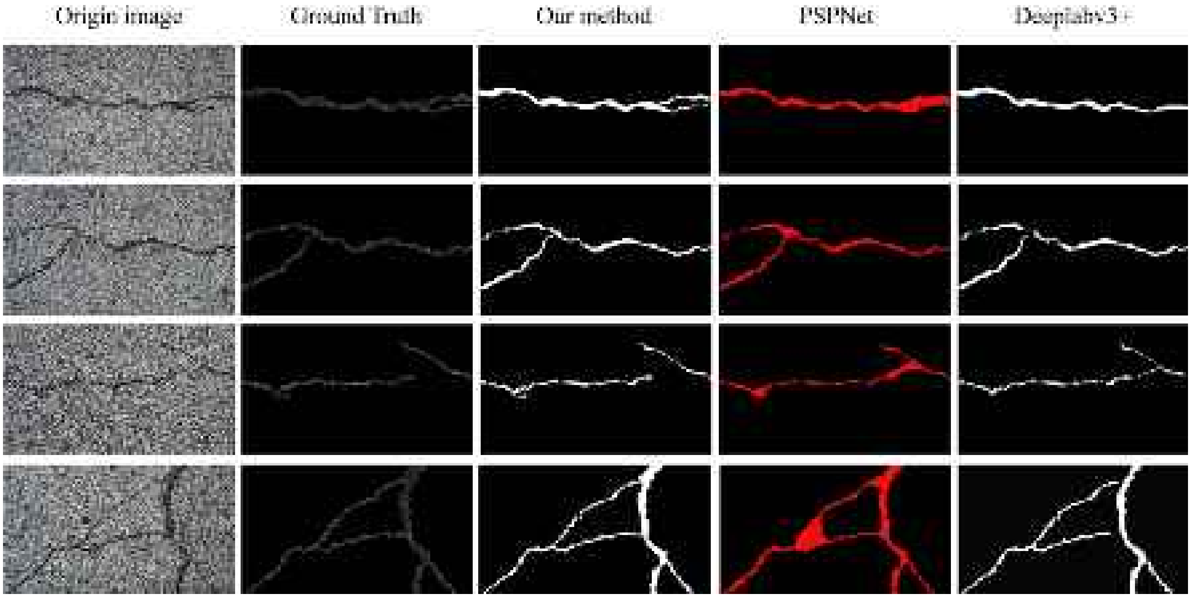

Fig. 5. Comparison of the segmentation results for our method, PSPNet, and Deeplabv3+.

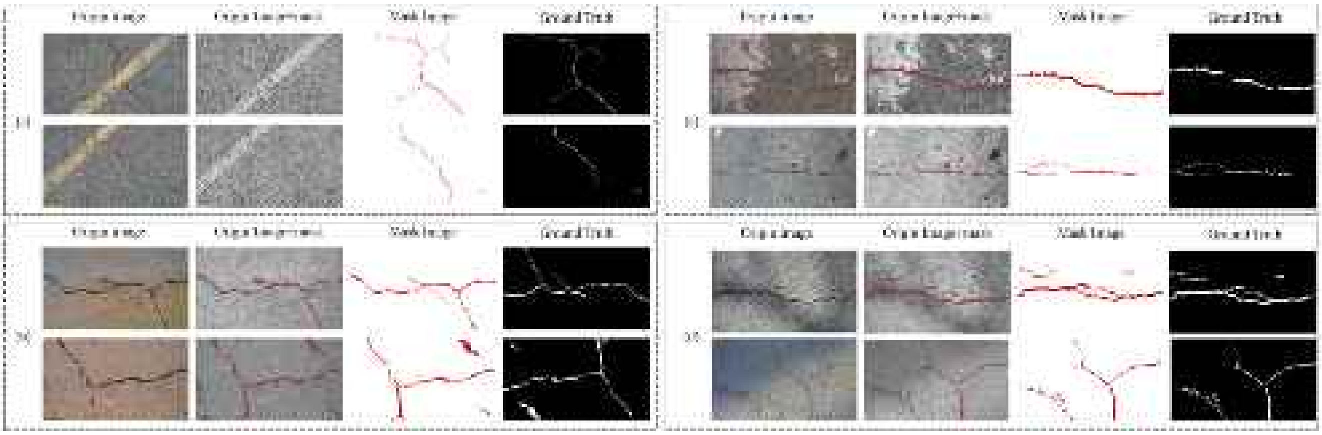

Fig. 6. Our method evaluates model detection results in complex detection scenarios, (a) foreign body fusion background, (b) aging background, (c) coloring background, and (d) light fusion background.

Table 6
Performance comparison of different improved modules on the Crack500 dataset.

Group	U-Net [9]	Attention gating	Softmax loss	Recall (%)	F1-Score (%)	MIoU (%)	FWIoU (%)	PA (%)
M1	✓			83.2	82.2	73.4	96.4	96.4
M2	✓	✓		83.6	83.6	74.6	96.6	97.5
M3	✓		✓	90.4	88.4	75.2	97.2	96.8
M4	✓	✓	✓	90.6	89.2	75.6	97.5	97.0

4.5. Ablation experiment

To further validate the effectiveness of different modules within AGSAM-Net, we conducted ablation experiments by incrementally adding the Attention Gating module and the Softmax Loss function. The evaluation metrics include Recall (%), F1-Score (%), MIoU (%), FWIoU (%), and PA (%). The results are summarized in Table 6.

Table 7
Performance comparison of different semantic segmentation networks on UAV generalization dataset.

Method	Recall (%)	F1-Score (%)	MIoU (%)	FWIoU (%)	PA (%)	MPA (%)
U-Net [9]	81.7	81.1	71.4	94.3	96.4	79.8
FCN [41]	83.8	82.3	70.3	94.9	97.2	80.2
SegNet [40]	81.3	80.9	72.8	95.6	95.2	79.8
Deeplabv3+ [42]	82.8	82.9	73.6	94.3	95.7	86.9
PSPNet [43]	86.6	83.9	73.5	94.5	97.4	83.7
AGSAM-Net	86.5	86.6	75.5	95.4	98.4	84.3

Table 8
CPA (%) comparison of bridge defect categories in different semantic segmentation networks.

Type	U-Net [9]	FCN [41]	SegNet [40]	Deeplabv3+ [42]	PSPNet [43]	AGSAM-Net
Honeycomb	66.3	72.1	67.5	74.2	74.6	75.3
Crack	64.1	66.7	63.2	68.8	73.2	74.5
Seepage	71.4	68.5	68.6	79.5	82.1	80.2
Repaired	84.4	75.9	68.4	77.8	79.2	80.6
Spalling	89.6	65.7	84.6	91.4	92.3	91.8
Others	94.2	95.6	95.8	96.4	97.8	99.4

As shown in Table 6, adding the Attention Gating module and the Softmax Loss function significantly enhances the model's performance. For example, the Recall metric increases from 83.2% in M1 to 90.6% in M4, an improvement of 7.4%. Similarly, the F1-Score improves by 7.0% (from 82.2% to 89.2%), and MIoU increases by 2.2% (from 73.4% to 75.6%). Furthermore, FWIoU and PA also show notable improvements, with FWIoU rising from 96.4% to 97.5% (a 1.1% increase) and PA improving from 96.4% to 97.0% (a 0.6% increase). The M4 group demonstrates the best overall performance across all metrics, particularly in Recall and F1-Score, showcasing AGSAM-Net's strong semantic segmentation capabilities. These results indicate that integrating the Attention Gating module and Softmax Loss function into U-Net effectively enhances the model's generalization and accuracy in semantic segmentation tasks. This improvement is particularly significant for the Crack500 dataset, demonstrating the practical advantages of this approach in real-world applications.

Moreover, the results reveal that the M3 group, which incorporates the Softmax Loss function, outperforms the M2 group that includes the Attention Gating module. This suggests that the Softmax Loss function is more effective in addressing the performance issues arising from the imbalance between positive and negative samples, thereby significantly improving the model's accuracy and robustness.

4.6. Generalization experiment

Fig. 6 illustrates the effectiveness of the proposed defect recognition method when applied to real-world bridge structure images. As shown in the figure, the method demonstrates strong segmentation capabilities across various challenging backgrounds, including surfaces with foreign object interference, aged surfaces, colored backgrounds, and surfaces with complex lighting conditions. Even in scenarios with extremely low or inadequate lighting, the method accurately identifies and segments the geometric shapes of cracks, highlighting its robust generalization ability and adaptability. This indicates that the crack recognition method based on AGSAM-Net offers superior environmental adaptability compared to other deep learning approaches. Furthermore, the evaluation metrics for bridge defect segmentation across different complex backgrounds are consistently higher than those of the comparison methods, underscoring the improved recognition performance of the proposed approach.

This round of experiments aims to compare the performance of several mainstream semantic segmentation networks on a generalized bridge surface defect image dataset. The evaluated networks include FCN, SegNet, U-Net, and Deeplabv3+, all of which use binary cross-entropy loss, while AGSAM-Net employs the Softmax Loss function. The evaluation metrics include overall Recall (%), F1-Score (%), MIoU (%), FWIoU (%), PA (%), and MPA (%). As shown in Table 7, AGSAM-Net demonstrates the best performance on the generalized bridge surface defect dataset. Specifically, AGSAM-Net achieves PA, MIoU, and FWIoU scores of 98.4%, 75.5%, and 95.4%, respectively, all of which surpass the other segmentation models. Although AGSAM-Net's MPA is slightly lower at 84.3% compared to Deeplabv3+'s 86.9%, AGSAM-Net still exhibits superior overall performance, particularly with an F1-Score of 86.6%, the highest among the five models. These results indicate that AGSAM-Net has high detection accuracy and practical value in real-world applications, significantly enhancing the practical capability of semantic segmentation while ensuring the recall rate in bridge surface defect detection. In the generalized bridge surface defect dataset of 2,800 test samples, AGSAM-Net accurately detected 1,322 defect-free image patches and 263 defective image patches, achieving an accuracy rate of 98.8%. These results further validate AGSAM-Net's robust performance in complex scenarios, making it an ideal choice for bridge surface defect detection tasks.

We evaluate the performance of the six semantic segmentation models across different defect categories. Table 8 presents the Category Pixel Accuracy (CPA) results for various defect types. Overall, AGSAM-Net exhibits the best performance. Specifically, AGSAM-Net achieves CPA values of 74.5%, 80.2%, and 91.8% in the challenging categories of cracks, irregular cracks, and spalling, respectively, outperforming the second-best model by approximately 1.3%, 0.7%, and 1.1%. In the seepage and spalling categories, AGSAM-Net's performance is slightly lower, with differences of only 0.4% and 0.5% compared to the best-performing Deeplabv3+ and PSPNet models. However, SegNet and U-Net show relatively poor performance, particularly in the seepage and repaired categories, indicating the need for further improvements and optimizations. Despite its slightly lower performance in the crack category, AGSAM-Net still demonstrates competitive CPA of 74.5%, highlighting its strong generalization capability in handling complex defect types.

Table 9 provides the Intersection over Union (IoU) results for different defect categories. AGSAM-Net excels in most categories, especially in irregular cracks, seepage, and repaired categories, where it achieves IoU values of 73.5%, 78.9%, and 81.9%, respectively, surpassing the second-best models by 1.3%, 1.1%, and 2.7%. In the honeycomb and spalling categories, the performance gap between AGSAM-Net and the top-performing PSPNet and Deeplabv3+ models is only 0.9% and 0.6%, respectively. These results demonstrate AGSAM-Net's superior generalization and detection accuracy across a variety of complex bridge surface defects.

By combining the CPA data in Table 8 with the IoU data in Table 9, it is evident that despite slightly lower performance in the crack and seepage categories, AGSAM-Net consistently maintains high accuracy across the overall detection tasks. Specifically, the model exhibits high sensitivity to large, concentrated defects such as repaired and

Table 9
IoU (%) comparison of bridge defect categories in different semantic segmentation networks.

Type	U-Net [9]	FCN [41]	SegNet [40]	Deeplabv3+ [42]	PSPNet [43]	AGSAM-Net
Honeycomb	65.9	70.8	68.5	74.0	74.5	75.6
Crack	65.5	67.8	62.3	68.2	74.1	73.5
Seepage	72.1	68.8	67.6	79.8	81.2	78.9
Repaired	84.7	76.5	67.5	76.7	79.2	81.9
Spalling	88.6	64.3	84.0	90.8	92.6	93.2
Others	93.2	97.0	95.9	96.0	96.4	99.6

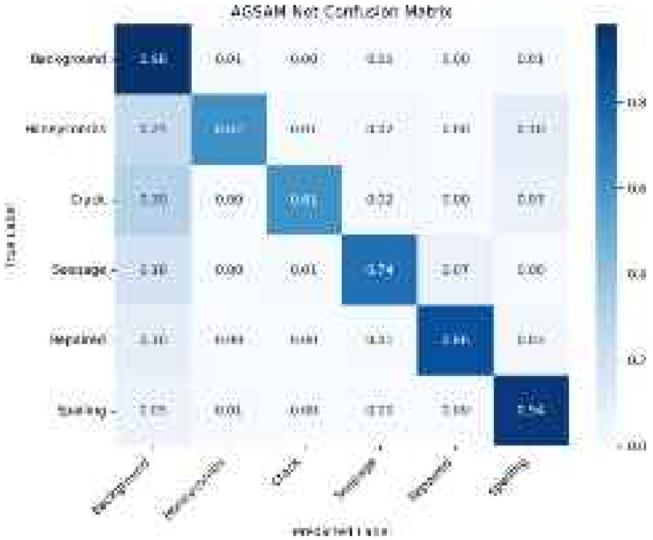

Fig. 7. Multiclassification confusion matrix for AGSAM-Net.

spalling categories, while detection performance is relatively weaker for smaller, irregular cracks and scattered defects. This may be due to the loss of contextual information caused by image tiling, which reduces classification effectiveness. Fig. 7 shows the multi-class confusion matrix of AGSAM-Net as supplementary evidence for this analysis.

5. Conclusion and discussion

In this study, we aimed to address the challenges of accuracy and efficiency in detecting bridge surface defects, particularly in complex scenarios with multiple defect types. To this end, we developed a UAV-based bridge defect detection system centered around an improved multi-class semantic segmentation network, AGSAM-Net. This network integrates an attention gating mechanism and an atrous spatial pyramid pooling (ASPP) module, enhancing its ability to capture multi-scale information and improving defect detection accuracy in challenging environments. Through extensive experiments, we validated AGSAM-Net's superior performance across various bridge datasets, particularly excelling in crack detection and defect quantification. The results demonstrated that AGSAM-Net outperforms existing segmentation models in key metrics such as MIoU and Recall, highlighting its potential for real-world applications.

This study successfully designed and implemented a bridge defect detection system based on AGSAM-Net, achieving significant improvements in detection accuracy and processing efficiency. By leveraging high-resolution images captured by UAVs and integrating them with the advanced segmentation capabilities of AGSAM-Net, the system accurately classifies and quantifies multiple bridge surface defects. The experimental results confirm that AGSAM-Net is highly robust in complex scenarios and effectively detects small cracks. By integrating multi-source data, the system provides strong technical support for bridge health monitoring and maintenance decision-making.

Despite AGSAM-Net's outstanding performance, certain limitations remain. While the network performs well in crack detection, it can still be prone to false positives when dealing with noise and complex backgrounds. Additionally, although the system successfully classifies various defect types, its accuracy in identifying rare or small-sample defects could be further improved. Future research could focus on incorporating more diverse datasets and more sophisticated feature fusion mechanisms to further enhance the model's accuracy and robustness.

Looking ahead, future research directions will focus on two primary areas: first, exploring the integration of the system with an embodied robotic arm to enable automated welding and repair of bridge defects; second, further optimizing the structure of AGSAM-Net to improve its real-time detection performance. Additionally, with advancements in hardware, achieving more efficient and precise UAV path planning and visual guidance will become a key area of exploration. We believe that through continuous technological innovation, this system will play a vital role in the intelligent maintenance of bridges.

CRediT authorship contribution statement

Rongji Li: Writing – original draft, Software, Conceptualization. **Ziqian Wang:** Writing – review & editing, Validation, Software, Formal analysis.

Declaration of competing interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Data availability

Data will be made available on request.

References

[1] F. Potenza, C. Rinaldi, E. Ottaviano, V. Gattulli, A robotics and computer-aided procedure for defect evaluation in bridge inspection, J. Civ. Struct. Heal. Monit. 10 (2020) 471–484.

[2] R. Pang, Y. Yang, A. Huang, Y. Liu, P. Zhang, G. Tang, Multi-scale feature fusion model for bridge appearance defect detection, Big Data Min. Anal. 7 (1) (2023) 1–11.

[3] J. Chen, Y. Wen, Y.A. Nanehkaran, D. Zhang, A. Zeb, Multiscale attention networks for pavement defect detection, IEEE Trans. Instrum. Meas. 72 (2023) 1–12.

[4] C. Xiong, T. Zayed, E.M. Abdelkader, A novel YOLOv8-GAM-wise-iou model for automated detection of bridge surface cracks, Constr. Build. Mater. 414 (2024) 135025.

[5] H. Sun, L. Song, Z. Yu, A deep learning-based bridge damage detection and localization method, Mech. Syst. Signal Process. 193 (2023) 110277.

[6] R. Li, J. Yu, F. Li, R. Yang, Y. Wang, Z. Peng, Automatic bridge crack detection using unmanned aerial vehicle and faster R-CNN, Constr. Build. Mater. 362 (2023) 129659.

[7] M. Hao, Z. Zhang, L. Li, K. Dong, L. Cheng, P. Tiwari, X. Ning, Coarse to fine-based image–point cloud fusion network for 3D object detection, Inf. Fusion 112 (2024) 102551.

[8] M. Zheng, Z. Lei, K. Zhang, Intelligent detection of building cracks based on deep learning, Image Vis. Comput. 103 (2020) 103987.

[9] Z. Liu, Y. Cao, Y. Wang, W. Wang, Computer vision-based concrete crack detection using U-net fully convolutional networks, Autom. Constr. 104 (2019) 129–139.

[10] T. Chen, Z. Cai, X. Zhao, C. Chen, X. Liang, T. Zou, P. Wang, Pavement crack detection and recognition using the architecture of segNet, J. Ind. Inf. Integr. 18 (2020) 100144.

[11] H. Fu, D. Meng, W. Li, Y. Wang, Bridge crack semantic segmentation based on improved Deeplabv3+, J. Mar. Sci. Eng. 9 (6) (2021) 671.

[12] Z. Li, H. Zhu, M. Huang, A deep learning-based fine crack segmentation network on full-scale steel bridge images with complicated backgrounds, IEEE Access 9 (2021) 114989–114997.

[13] Q. Chen, F. He, G. Wang, X. Bai, L. Cheng, X. Ning, Dual guidance enabled fuzzy inference for enhanced fine-grained recognition, IEEE Trans. Fuzzy Syst. (2024) 1–14.

[14] A.M. Alani, M. Aboutalebi, G. Kilic, Applications of Ground Penetrating Radar (GPR) in bridge deck monitoring and assessment, J. Appl. Geophys. 97 (2013) 45–54.

[15] M. Rashidi, M. Mohammadi, S. Sadeghlou Kivi, M.M. Abdolvand, L. Truong-Hong, B. Samali, A decade of modern bridge monitoring using terrestrial laser scanning: Review and future directions, Remote Sens. 12 (22) (2020) 3796.

[16] H. Zhao, Y. Ding, A. Li, W. Sheng, F. Geng, Digital modeling on the nonlinear mapping between multi-source monitoring data of in-service bridges, Struct. Control Health Monit. 27 (11) (2020) e2618.

[17] A. Ellenberg, A. Kontsos, F. Moon, I. Bartoli, Bridge deck delamination identification from unmanned aerial vehicle infrared imagery, Autom. Constr. 72 (2016) 155–165.

[18] T. Song, X. Zhang, D. Yang, Y. Ye, C. Liu, J. Zhou, Y. Song, Lightweight detection network based on receptive-field feature enhancement convolution and three dimensions attention for images captured by UAVs, Image Vis. Comput. 140 (2023) 104855.

[19] 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? Constr. Build. Mater. 347 (2022) 128543.

[20] T. Omar, M.L. Nehdi, Remote sensing of concrete bridge decks using unmanned aerial vehicle infrared thermography, Autom. Constr. 83 (2017) 360–371.

[21] Z. Mu, Y. Qin, C. Yu, Y. Wu, Z. Wang, H. Yang, Y. Huang, Adaptive cropping shallow attention network for defect detection of bridge girder steel using unmanned aerial vehicle images, J. Zhejiang Univ.-Science A 24 (3) (2023) 243–256.

[22] B.J. Perry, Y. Guo, R. Atadero, J.W. van de Lindt, Streamlined bridge inspection system utilizing Unmanned Aerial Vehicles (UAVs) and machine learning, Measurement 164 (2020) 108048.

[23] J.J. Lin, A. Ibrahim, S. Sarwade, M. Golparvar-Fard, Bridge inspection with aerial robots: Automating the entire pipeline of visual data capture, 3D mapping, defect detection, analysis, and reporting, J. Comput. Civ. Eng. 35 (2) (2021) 04020064.

[24] O. Avci, O. Abdeljaber, S. Kiranyaz, M. Hussein, M. Gabbouj, D.J. Inman, A review of vibration-based damage detection in civil structures: From traditional methods to machine learning and deep learning applications, Mech. Syst. Signal Process. 147 (2021) 107077.

[25] F. Ni, J. Zhang, M.N. Noori, Deep learning for data anomaly detection and data compression of a long-span suspension bridge, Comput.-Aided Civ. Infrastruct. Eng. 35 (7) (2020) 685–700.

[26] S. Dorafshan, H. Azari, Deep learning models for bridge deck evaluation using impact echo, Constr. Build. Mater. 263 (2020) 120109.

[27] D. Isailović, V. Stojanovic, M. Trapp, R. Richter, R. Hajdin, J. Döllner, Bridge damage: Detection, IFC-based semantic enrichment and visualization, Autom. Constr. 112 (2020) 103088.

[28] X. Jin, M.Z. Haider, Y. Cui, J.G. Jang, Y.J. Kim, G. Fang, J.W. Hu, Development of nanomodified self-healing mortar and a U-Net model based on semantic segmentation for crack detection and evaluation, Constr. Build. Mater. 365 (2023) 129985.

[29] H. Yuan, T. Jin, X. Ye, Modification and evaluation of attention-based deep neural network for structural crack detection, Sensors 23 (14) (2023) 6295.

[30] D. Amirkhani, M.S. Allili, L. Hebbache, N. Hammouche, J.-F. Lapointe, Visual concrete bridge defect classification and detection using deep learning: A systematic review, IEEE Trans. Intell. Transp. Syst. (2024).

[31] S. Jiang, J. Zhang, W. Wang, Y. Wang, Automatic inspection of bridge bolts using unmanned aerial vision and adaptive scale unification-based deep learning, Remote Sens. 15 (2) (2023) 328.

[32] J. Guo, P. Liu, B. Xiao, L. Deng, Q. Wang, Surface defect detection of civil structures using images: Review from data perspective, Autom. Constr. 158 (2024) 105186.

[33] H. Wan, L. Gao, Z. Yuan, H. Qu, Q. Sun, H. Cheng, R. Wang, A novel transformer model for surface damage detection and cognition of concrete bridges, Expert Syst. Appl. 213 (2023) 119019.

[34] B. Chen, H. Zhang, G. Wang, J. Huo, Y. Li, L. Li, Automatic concrete infrastructure crack semantic segmentation using deep learning, Autom. Constr. 152 (2023) 104950.

[35] K. Luo, X. Kong, J. Zhang, J. Hu, J. Li, H. Tang, Computer vision-based bridge inspection and monitoring: A review, Sensors 23 (18) (2023) 7863.

[36] A. Sullivan, X. Lu, ASPP: a new family of oncogenes and tumour suppressor genes, Br. J. Cancer 96 (2) (2007) 196–200.

[37] W. Deng, Y. Mou, T. Kashiwa, S. Escalera, K. Nagai, K. Nakayama, Y. Matsuo, H. Prendinger, Vision based pixel-level bridge structural damage detection using a link ASPP network, Autom. Constr. 110 (2020) 102973.

[38] C. Zhang, C.-c. Chang, M. Jamshidi, Concrete bridge surface damage detection using a single-stage detector, Comput.-Aided Civ. Infrastruct. Eng. 35 (4) (2020) 389–409.

[39] F. Yang, L. Zhang, S. Yu, D. Prokhorov, X. Mei, H. Ling, Feature pyramid and hierarchical boosting network for pavement crack detection, IEEE Trans. Intell. Transp. Syst. 21 (4) (2019) 1525–1535.

[40] V. Badrinarayanan, A. Kendall, R. Cipolla, Segnet: A deep convolutional encoder-decoder architecture for image segmentation, IEEE Trans. Pattern Anal. Mach. Intell. 39 (12) (2017) 2481–2495.

[41] Y. Liu, J. Yao, X. Lu, R. Xie, L. Li, DeepCrack: A deep hierarchical feature learning architecture for crack segmentation, Neurocomputing 338 (2019) 139–153.

[42] L.-C. Chen, Y. Zhu, G. Papandreou, F. Schroff, H. Adam, Encoder-decoder with atrous separable convolution for semantic image segmentation, in: Proceedings of the European Conference on Computer Vision, ECCV, 2018, pp. 801–818.

[43] J. Zhou, M. Hao, D. Zhang, P. Zou, W. Zhang, Fusion pspnet image segmentation based method for multi-focus image fusion, IEEE Photon. J. 11 (6) (2019) 1–12.