13th International Conference on
Structural Health Monitoring of Intelligent Infrastructure
DOI: 10.3217/978-3-99161-057-1-049

Universal unsupervised image segmentation model of multi-type component and damage for vision-based autonomous UAV inspection of bridges

Guangshuo YANG[1], Chuao ZHANG[1], Yang XU[1,2*] 0000-0002-8394-9224
[1]School of Civil Engineering, Harbin Institute of Technology, Harbin, 150090, China
[2]Key Lab of Smart Prevention and Mitigation of Civil Engineering Disasters of the Ministry of Industry and Information Technology, Harbin 150090, China
*Corresponding Author: Dr. Yang XU, xyce@hit.edu.cn

ABSTRACT: Although recent advances have been widely gained in UAV-based visual inspection for bridges, the accuracy and generalization ability of recognition model highly rely on sufficient, complete, and high-quality annotations. Current damage segmentation models are often trained in a fragmented manner based on substantial pixel-level labels for specific structural components and damage types, lacking universality and robustness under real-world open scenarios. This study establishes a universal unsupervised image segmentation model of multi-type component and damage for vision-based autonomous UAV inspection of bridges using a teacher-student network architecture. The inputs are unlabeled image pairs after data augmentation including random clipping, rotation, illumination transformation, and color transformation. The pre-trained backbone of original DINO is adopted as frozen image feature extractor to obtain high-level feature representations, and a CNN-based segmentation head with learnable parameters is designed to generate dense segmentation maps with strong point-wise correlations. A synthetic loss function, comprising a correlation loss and a contrastive loss, is proposed for model training. The proposed method is validated on a unified multi-scale imageset including various structural components and surface damage for cable-supported bridges and concrete bridges. The recognition accuracy, generalization ability, and robustness under complex background are demonstrated.

KEY WORDS: UAV Bridge Inspection; Universal Unsupervised Segmentation; Teacher-Student Network; Cross-level Feature Alignment; Contrastive Learning.

1 INTRODUCTION

Maintaining safe operation throughout the entire life cycle of bridge structures is crucial. Timely and accurate identification of surface damage (such as cracks, corrosion, etc.) not only provides a basis for scientifically formulating maintenance strategies but also effectively prevents structural performance degradation, significantly reducing major safety risks like collapse and instability. For decades, bridge inspection has primarily relied on manual methods. However, this approach is not only time-consuming and labor-intensive but also susceptible to inspector experience, environmental conditions, and fatigue factors, resulting in issues such as strong subjectivity, low efficiency, and poor consistency, making it difficult to ensure reliability and timeliness in practical applications [1].

In recent years, computer vision technology has demonstrated significant advantages in image-based structural health monitoring and damage identification, providing an efficient and reliable alternative to traditional manual inspection methods. By integrating digital image processing with machine learning algorithms, computer vision enables automated identification and quantitative analysis of structural surface damage. Early research mainly focused on traditional image processing methods like edge detection and threshold segmentation, but their performance heavily relied on manual parameter tuning and showed limited generalization capability for damage features in complex environments. With breakthroughs in deep learning, data-driven methods represented by convolutional neural networks (CNNs) have exhibited outstanding performance in automatic feature extraction and damage pattern recognition, greatly improving the accuracy and adaptability of structural health monitoring systems. However, these methods still depend on manually designed features as input and face challenges in robustness across complex real-world scenarios and generalization across different damage types [2].

Deep learning achieves end-to-end automatic mapping between images and object annotations through deep neural networks, with CNNs as multi-level feature extractors being the most extensively studied [3-5]. Existing structural damage identification methods are typically developed based on specific datasets covering only limited damage types and application scenarios, resulting in constrained generalization capability for new damage categories or under disaster conditions [6-7]. Moreover, these methods often require large amounts of annotated data to achieve ideal performance, but the nonlinear and sparse nature of structural damage makes high-quality annotated data difficult to obtain in practice [8-9]. Consequently, identification accuracy is easily affected by sample size, class balance, and damage diversity. However, real-world engineering applications demand models that maintain good generalization across diverse scenarios while achieving high-precision identification performance on unannotated data [10-11]. To overcome the limitations of existing methods trained separately on different datasets, it is necessary to develop a universal visual recognition model for structural damage that can accurately identify multiple damage types while maintaining stable performance in complex backgrounds and multi-scale real-world scenarios [12].

In the field of structural health monitoring, unsupervised and self-supervised learning paradigms are gradually becoming key technological pathways to address few-shot damage detection

DOI: 10.3217/978-3-99161-057-1-049

challenges [13-14]. Currently, although contrastive learning-driven unsupervised semantic segmentation methods have shown potential in general computer vision domains, their adaptation to bridge UAV visual inspection scenarios faces significant bottlenecks: on one hand, there is a need to develop a universal segmentation framework adapted to Transformer architectures; on the other hand, the challenge of pixel-level damage parsing in real-world environments with massive unannotated data must be overcome (although certain few-shot learning algorithms have been proposed [15-17], a proper number of samples with pixel-level annotations should be required in advance). To address these challenges, this study proposes a large vision model for universal structural damage segmentation. The main scientific contributions are summarized as follows:

(1) We propose an unsupervised structural damage segmentation method based on teacher-student network knowledge distillation, using unlabeled augmented image pairs as input. The dual-branch architecture incorporates pre-trained frozen Transformer backbones and fine-tunable CNN segmentation heads.

(2) We design a dual-strategy collaborative mechanism combining cross-level self-supervised correlation learning and cross-network contrastive learning. The former achieves feature alignment through point-to-point correlations between high-level features and dense segmentation maps, while the latter maintains instance similarity and separability through feature vector comparison between teacher-student networks.

(3) We construct a joint optimization objective integrating correlation loss and contrastive loss. The student branch is rapidly updated via gradient descent, while the teacher branch is stably adapted through momentum-based exponential moving average, achieving end-to-end fine-tuning of the segmentation head.

The remainder of this paper is organized as follows. Section 2 introduces the network architecture of the proposed universal unsupervised damage segmentation model. Section 3 describes the investigated imageset of multi-scale multi-type structural components and surface damage. Section 4 presents a series of test results to demonstrate the effectiveness, robustness, and generalization capability of the established model under real-world inspection scenarios with complex background disturbances for cable-supported bridges and concrete bridges. Finally, Section 5 concludes this paper.

2 METHODOLOGY

The architecture of the proposed universal structural damage segmentation model is illustrated in Figure 1, employing an unsupervised learning paradigm based on collaborative optimization of teacher-student networks through an end-to-end self-supervised knowledge distillation mechanism for effective feature learning. Distinct from conventional approaches dependent on manual annotations, our framework integrates three fundamental components: a data augmentation module, a Transformer-based frozen feature extraction backbone, and a tunable CNN segmentation head. A specially designed hybrid loss function combining feature-space correlation loss with instance contrastive loss enables the network to autonomously identify essential damage characteristics in unsupervised settings. During deployment, input images undergo sequential processing through the frozen feature extraction backbone, optimized segmentation head, and semantic clustering-based post-processing module to generate pixel-accurate structural damage segmentation results, with the complete pipeline demonstrating enhanced robustness for practical engineering applications. This integrated approach effectively bridges the gap between unsupervised learning and precise damage identification in complex real-world scenarios while maintaining computational efficiency throughout the segmentation process.

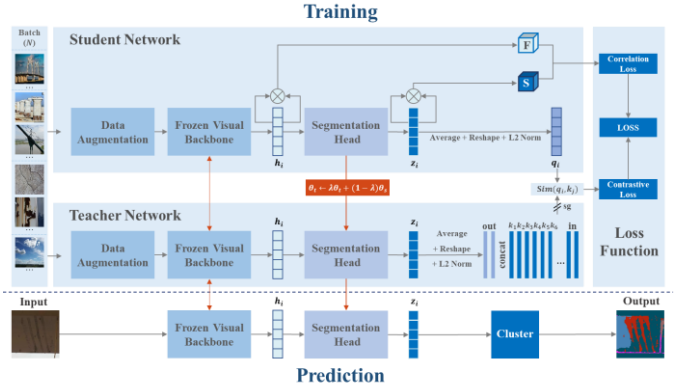


Figure 1. Model architecture for universal unsupervised segmentation of multi-type structural component and damage.

For each individual instance of input image, feature maps with the same dimensions of channel, height, and width are obtained before and after the segmentation head. For each branch of student and teacher networks, spatial points on feature maps before the segmentation head are noted as f_{chw} and $g_{ch'w'}$, while spatial points on feature maps after the segmentation head are noted as s_{chw} and $t_{ch'w'}$, respectively. The feature correspondence $F_{hwh'w'}$ between f_{chw} and $g_{ch'w'}$ and segmentation correspondence $S_{hwh'w'}$ between s_{chw} and $t_{ch'w'}$ are obtained by calculating the point-wise cosine similarity as

$$F_{hwh'w'} = \frac{\sum_{c=1}^{C} f_{chw} \times g_{ch'w'}}{\|f_{hw}\|_2 \times \|g_{h'w'}\|_2} \quad (1)$$

$$S_{hwh'w'} = \frac{\sum_{c=1}^{C} s_{chw} \times t_{ch'w'}}{\|s_{hw}\|_2 \times \|t_{h'w'}\|_2} \quad (2)$$

For N input images within an input batch, the feature correspondence tensors and segmentation correspondence tensors could be denoted as $F_1, \ldots, F_N \in \mathcal{R}^{H\times W\times H\times W}; S_1, \ldots, S_N \in \mathcal{R}^{H\times W\times H\times W}$ with four-dimensional elements of $F_{hwh'w'}$ and $S_{hwh'w'}$.

The dense semantic correlation loss L_corr is calculated based on feature correspondence tensors and segmentation correspondence tensors by

$$F^{SC}_{hwh'w'} = F_{hwh'w'} - \frac{1}{HW}\sum_{h',w'} F_{hwh'w'} \quad (3)$$

$$L_{corr} = -\sum_{h,w,h',w'} (F^{SC}_{hwh'w'} - b) max(S_{hwh'w'}, 0) \quad (4)$$

where $F^{SC}_{hwh'w'}$ denotes the feature correspondence tensor after spatial centralization, b is a hyperparameter to avoid model collapse and ensure a positive correlation loss value for loss descending.

DOI: 10.3217/978-3-99161-057-1-049

The contrastive loss between the teacher-student networks is defined as

$$L_{cont} = -\sum_{i=1}^{N} log\left\{\frac{exp[Sim(q_i,k_+)/\tau]}{\sum_{j=1}^{K} exp[Sim(q_i,k_j)/\tau]}\right\} \quad (5)$$

where Sim denotes the cosine similarity between two vectors, q_i denotes the ith query feature vector obtained from the student branch for the ith image in a batch, k_+ denotes the feature vector obtained from the teacher branch as the positive sample of the corresponding query image, N denotes the batch size, k_j denotes the jth referenced feature vector in the feateure dictionary, K denotes the queue length of the preset feateure dictionary, τ denotes a temperature hyperparameter to enhance an exponential amplification effect.

The synthetic loss function is defined by a weighted sum of correlation loss and contrastive loss as

$$Loss = \alpha L_{corr} + (1-\alpha)L_{cont} \quad (6)$$

where α denotes the weight coefficient of the correlation loss.

Upon completion of training, the system can directly generate predicted segmentation results for new test images using the frozen visual backbone network and optimized segmentation head. As shown in Figure 2, the prediction network incorporates a semantic clustering post-processing module consisting of K-means clustering and fully connected Conditional Random Field (CRF). Based on predefined label-category mapping relationships, the system automatically associates pixel-level annotations with specific categories including structural components, surface damage, and background. For newly added categories, only a single annotated sample is required to update the label mapping table. It should be noted that as standard post-processing modules for unsupervised image segmentation, the specific implementations of K-means clustering, fully connected CRF, and label alignment can be referenced in existing literature. To highlight this study's core contribution - the construction of a large-scale model for universal structural damage segmentation - the relevant implementation details are omitted here for brevity.

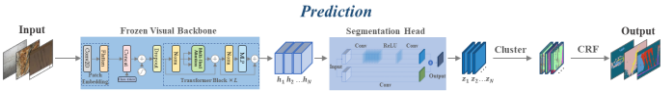


Figure 2. Schematic of prediction network structure with post-processing of semantic clustering.

3 IMPLEMENTATION DETAILS

This study addresses the key scientific challenge of unsupervised semantic segmentation for multiple types of structural damage under complex engineering conditions, proposing an integrated analytical framework that combines multi-source environmental interference factors with intrinsic structural features. Specifically, by fusing multi-dimensional key characteristics—including environmental background noise interference, macro-structural morphological features, micro-component texture details, and spatial distribution patterns of cable systems—we developed a dedicated scene model for damage detection in typical bridge structures such as cable-stayed bridges and concrete bridges. As illustrated in Figure 3, representative damage samples from the two-level structural damage image database constructed in this study are presented.

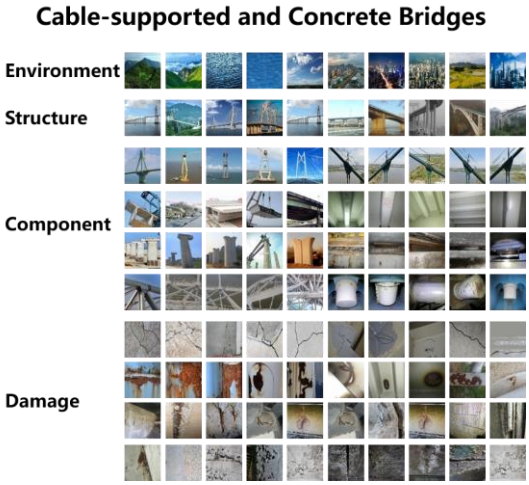


Figure 3. Representative images of hierarchical structural damage with multi-scale information for cable-supported and concrete bridges.

A total of 20K images with varying resolutions were standardized to a uniform resolution of 1,024 × 1,024 pixels. Each resized image was subsequently partitioned into 224 × 224 patches using a sliding window approach with a 100-pixel stride, thereby generating an extensive collection of image patches while circumventing potential feature degradation associated with direct downsampling. From this collection, 128 patches were randomly sampled to constitute the input batches for the proposed methodology.

It must be particularly emphasized that implementing traditional supervised CNN-based semantic segmentation models on such large-scale image patches inevitably encounters multiple computational challenges. Specifically, the exponential growth in computational complexity caused by high-resolution input space, coupled with the enormous manual annotation costs and time resources required for pixel-level labeling, constitutes two fundamental bottlenecks. These core limitations fundamentally undermine the feasibility of traditional supervised learning paradigms in the current application scenario, thereby significantly diminishing their practical value.

Through extensive experimental validation and parameter tuning, this study has ultimately determined the hyperparameter configuration scheme for model training as shown in Table 1. It is particularly important to note that while the current parameter settings may not represent the global optimal solution, empirical research demonstrates that this configuration ensures the large vision model for general structural damage segmentation achieves satisfactory segmentation accuracy, maintains excellent robustness against complex background interference, and exhibits strong generalization capability for new scenarios. Based on this, the primary objective of this research is not to pursue the optimal combination of hyperparameters, but rather to systematically validate through experiments the technical feasibility and practical effectiveness of the proposed large vision model in achieving general structural damage segmentation in real-world detection scenarios.

DOI: 10.3217/978-3-99161-057-1-049

Table 1. Configurations of key training hyperparameters.

Hyperparameter Variables	Values
Number of images included in each batch	128
Dimension of feature representation after patch embedding	512
Length of query, key, and value vectors in attention mechanism	64
Number of multiple attention heads	8
Number of stacked transformer blocks	6
Dimensions of extracted feature maps by frozen visual backbone	16, 16, 8
Length of query and referenced feature vectors in preset feature dictionary	64
Number of feature queue size	12,800
Positive parameter to avoid model collapse in correlation loss	0.18
Weight coefficient of correlation loss	0.67
Temperature coefficient of contrastive loss	0.07
Learning rate in stochastic gradient descent updating of student network	5e-4
Momentum in exponential moving average updating of teacher network	0.999
Number of training iterations	5,000

To obtain quantitative evaluation metrics for semantic segmentation, a set of test images are pre-labelled with pixel-level annotations, and the following pixel accuracy (PA), mean intersection-over-union (mIoU), and frequency-weighted intersection-over-union (FWIoU) are calculated by

$$\mathrm{PA} = \frac{\sum_i p_{ii}}{\sum_i \sum_j p_{ij}} \tag{7}$$

$$\mathrm{mIoU} = \frac{1}{N_C + 1} \sum_{i=0}^{N_C} \frac{p_{ii}}{\sum_{j \neq i} p_{ij} + \sum_{j \neq i} p_{ji} - p_{ii}} \tag{8}$$

$$\mathrm{FWIoU} = \frac{1}{\sum_{i=0}^{N_C} \sum_{j=0}^{N_C} p_{ij}} \sum_{i=0}^{N_C} \frac{p_{ii} \sum_{j=0}^{N_C} p_{ij}}{\sum_{j \neq i} p_{ij} + \sum_{j \neq i} p_{ji} - p_{ii}} \tag{9}$$

where p_{ij} denotes the number of pixels in the ith class (actual category) classified to the jth class (predicted category), the total number of pixel categories equals $N_C + 1$, including 1 for the background and N_C for the foreground, and N_C is determined by actual label categories of selected test images.

The proposed large vision model for universal structural damage segmentation is trained and tested under the software environment of PyTorch 1.8 and Python 3.7 on a 48G GPU of NVIDIA RTX A6000, and the average training time with the reported hyperparameter configurations is about 48 hours to obtain a well-trained model.

4 RESULTS AND DISCUSSION

Figure 4 shows some representative prediction results on coarse-grained segmentation of main bridge structures: (a) for cable-supported bridges and (b) for concrete bridges. The test PA, mIoU, and FWIoU are 97.17%, 91.47%, 94.64% for cable-support bridges and 92.72%, 82.46%, 86.98% for concrete bridges. The results show that main components of pylon, cable, girder, deck, and pier can be generally identified from entire images of bridge structures.

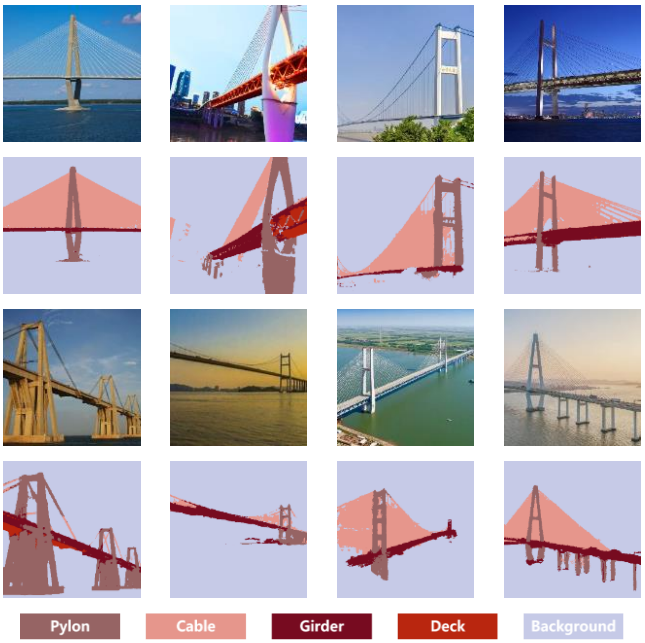
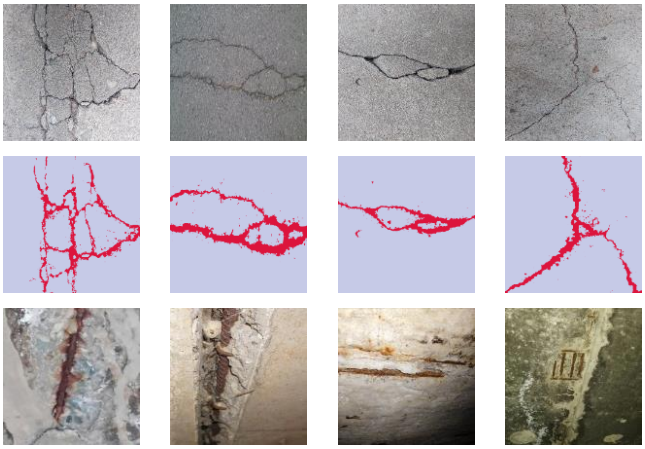


Figure 4. Representative predictions on coarse-grained segmentation of main bridge structures.

Figure 5 shows some typical prediction results of fine-grained damage segmentation on bridges, including concrete cracks, spalling, exposed rebar, seepage, salt damage, rebar fatigue cracks, coating peeling, and corrosion. Table 2 lists the evaluation metrics. The results show that this method can effectively detect various bridge damages from close-range images. Moreover, it can distinguish between combined concrete spalling and rebar exposure, as well as separate severe and slight corrosion areas.



DOI: 10.3217/978-3-99161-057-1-049

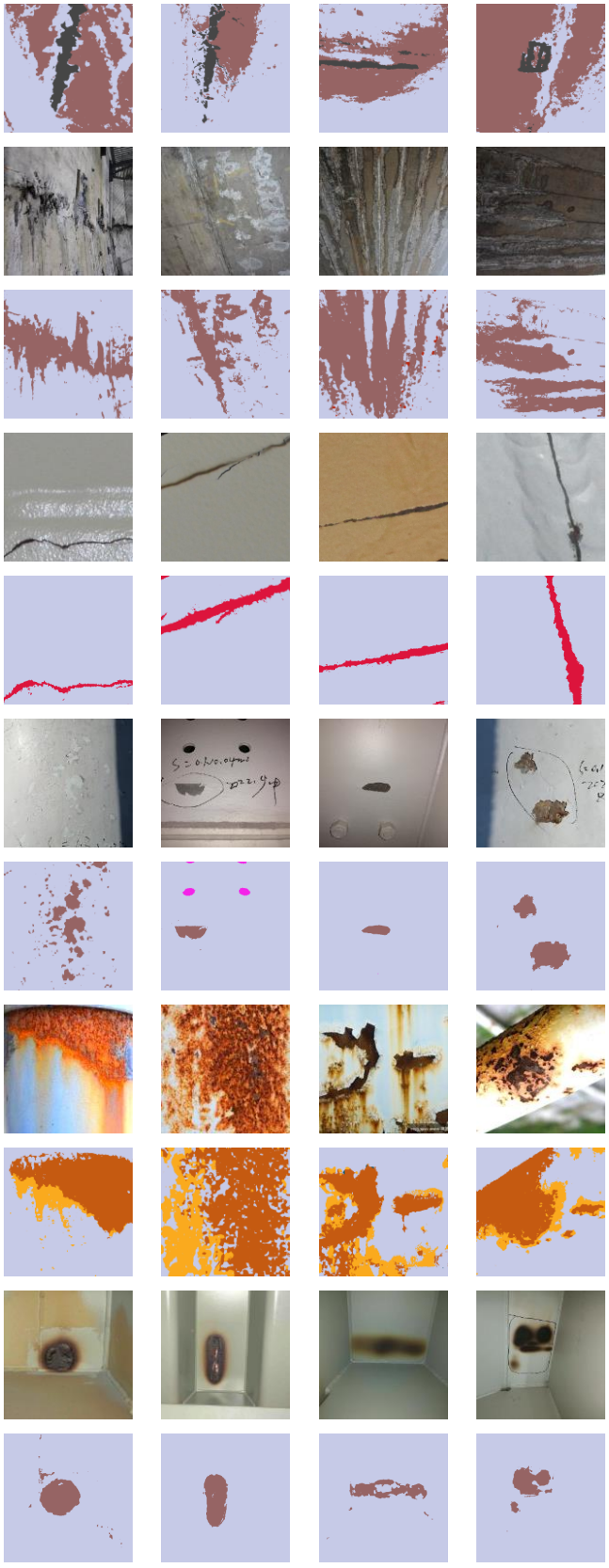
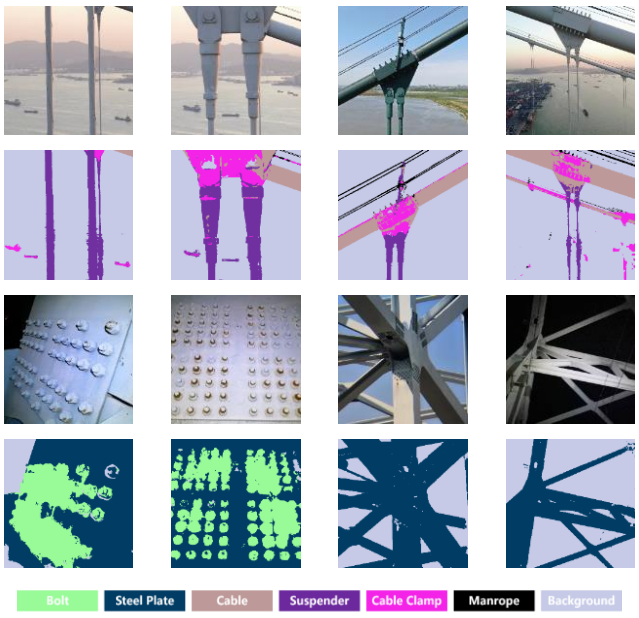


Figure 5. Representative predictions on fine-grained multi-type structural damage for bridges.

Table 2. Evaluation metrics for multi-type structural damage segmentation of bridges.

Bridge structural damage type	PA	mIoU	FWIoU
Concrete crack	96.19%	69.85%	93.35%
Concrete spalling/rebar exposure	98.97%	74.12%	98.30%
Water seepage/saltpetering	88.28%	75.15%	79.21%
Steel fatigue crack	96.21%	68.07%	94.69%
Coating spalling/steel corrosion	91.07%	75.39%	84.83%
Fire burning	95.85%	76.22%	93.02%

Figure 6 shows some representative prediction results on new bridge components and ship collision damage, and Table 3 shows the corresponding evaluation metrics. The results suggest that the proposed method have achieved generalization capacity on unseen categories of bridge components and damage apart from the existing structural components and surface damage included in the training imageset. Figure 12(a) shows that despite some misrecognitions of background boats with similar color to cable clamp, key components of cable-suspension bridges such as cable, suspender, cable clamp, and bolts on steel plate can be individually recognized. Figure 12(b) demonstrates that even for a never-appeared emergency of ship collision, the deck scratch, buckling and deformation of steel plate, coating spalling, and handrail failure fragments could be generally identified. It should be noted that some misrecognitions have been observed between the cable plane and distant background and that spurs-like pixels occur along the boundary of damage regions. Possible reasons might attribute to similar material and morphological features with indistinguishable member edges. Geometrical constraints of different component and damage regions would be further considered to address these issues in future study.



DOI: 10.3217/978-3-99161-057-1-049

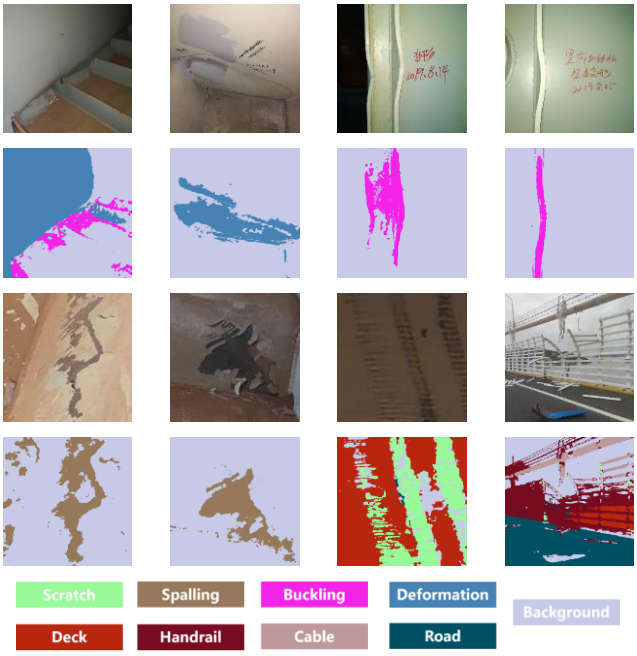


Figure 6. Representative predictions on new bridge components and ship collision damage.

Table 3. Evaluation metrics for segmentation of new bridge components.

New bridge component type	PA	mIoU	FWIoU
Bolt	81.39%	63.88%	70.97%
Steel plate	82.25%	69.24%	70.21%
Cable	96.23%	78.87%	93.84%
Suspender	96.70%	81.51%	94.37%
Cable clamp	95.19%	61.46%	90.96%

Furthermore, the effectiveness and necessity of the proposed method are demonstrated via performance comparisons between the proposed model and SCSegamba [18], a recently-reported supervised structural damage segmentation model. The model architecture of SCSegamba is shown in Figure 7. The core architecture incorporates a structure-aware visual state space (SAVSS) module and a multi-scale feature segmentation (MFS) head. The SAVSS captures continuous textures of multi-directional cracks through an innovative structure-aware scanning strategy and dynamically enhances crack features using a lightweight gated bottleneck convolution (GBC). Meanwhile, the MFS integrates multi-scale information to generate refined segmentation maps. Using a unified dataset (containing both concrete cracks and spalling/exposed reinforcement damages), we trained SCSegamba for 100 epochs using the default training setups and selected the optimal model (with minimal validation loss) for evaluation. Quantitative results demonstrate that our method outperforms SCSegamba across all three metrics (PA, mIoU, and FWIoU) for both crack and spalling/rebar exposure damage types (see detailed values in Table 4).

Figure 8 presents typical comparative results between the proposed method and SCSegamba in structural damage segmentation. The experimental results demonstrate that compared to the blurred boundaries and missed reinforcement detections generated by SCSegamba, our method can more accurately capture fine crack branches and complex boundary contours of spalling areas. Particularly in the identification tasks of concrete spalling and exposed reinforcement damage, the proposed teacher-student network architecture effectively enhances recognition robustness in weak-texture regions through the synergistic optimization mechanism of feature distillation and contrastive learning. Experimental results indicate that this method can precisely capture edge irregularities in spalling areas while maintaining high sensitivity to exposed reinforcement textures in low-contrast environments, ultimately achieving pixel-level precision in damage segmentation.

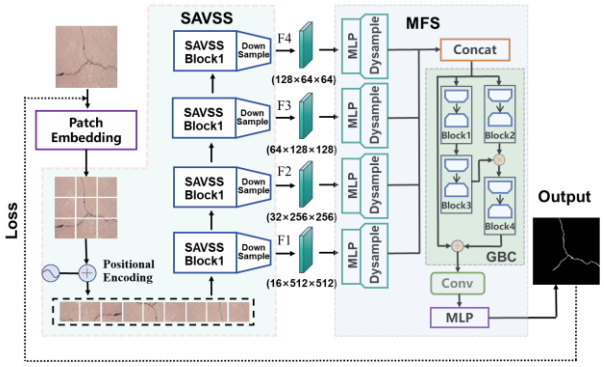


Figure 7. Model architecture of SCSegamba as comparative validation (reproduced from [18]).

Table 4. Performance comparison with supervised crack segmentation model SCSegamba on multi-class damage.

Method	Bridge structural damage type	PA	mIoU	FWIoU
SCSegamba [18]	Concrete crack	96.11%	65.20%	92.77%
Ours		96.19% (↑)	69.85% (↑)	93.35% (↑)
SCSegamba [18]	Concrete spalling/ rebar exposure	82.38%	60.47%	85.43%
Ours		95.19% (↑)	61.46% (↑)	90.96% (↑)

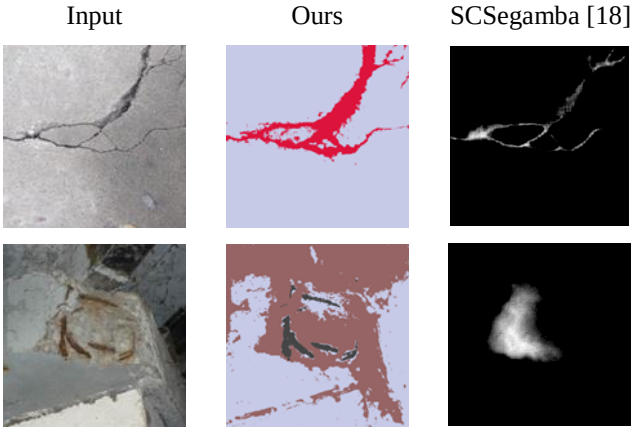


Figure 8. Some representative comparative results of structural damage segmentation between the proposed method and SCSegamba.

DOI: 10.3217/978-3-99161-057-1-049

5 CONCLUSIONS

This study proposes an unsupervised structural damage semantic segmentation framework based on relational learning and contrastive learning. By constructing a cross-scale feature interaction mechanism and a dynamic negative sample mining strategy, the framework achieves robust recognition and pixel-level localization of multiple types of structural damage in complex engineering scenarios without manual annotation. It effectively addresses the strong dependency of traditional supervised segmentation paradigms on large-scale fine-grained annotated data and the resulting limitations in model generalization. The main research conclusions are as follows:

(1) Based on a teacher-student network knowledge distillation framework, a unified semantic segmentation architecture for multiple types of structural components and surface damages was constructed. Each branch of the teacher-student network contains a pre-trained Transformer visual backbone and a fine-tunable CNN segmentation head.

(2) Using randomly augmented unlabeled image pairs as input, a pre-trained DINO backbone is employed as a frozen feature extractor to generate high-level feature maps. A CNN segmentation head with learnable parameters is designed to produce dense segmentation maps that maintain strong point-wise correlations with the high-level feature maps.

(3) Proposed an inter-layer correlation learning strategy between high-level feature maps from the frozen backbone network and dense segmentation maps from the fine-tuned segmentation head, achieving cross-level feature alignment for different structural components and damage regions within a single image. Developed a contrastive learning module with normalized feature aggregation between teacher-student branches to quantify intra-instance similarity and inter-instance discriminability across different images.

(4) A synthetic loss function comprising a correlation loss and a contrastive loss is designed. The segmentation head is efficiently fine-tuned by fastly optimizing the student network with direct error backpropagation by gradient descent and stably adapting the teacher network with exponential moving average by momentum updating.

(5) This study constructs a multi-scale image dataset encompassing various types of bridge structures and their damage patterns. Through systematic comparative experiments, the proposed model has demonstrated outstanding segmentation accuracy, strong generalization capability, and excellent robustness under complex background interference. The experimental results indicate that the large-scale visual model developed in this study has successfully achieved deep visual understanding of unlabeled bridge component damage images and effectively mastered their unsupervised learning mechanism.

ACKNOWLEDGMENTS

Financial support for this study was provided by the National Key R&D Program of China [Grant No. 2023YFC3805800], National Natural Science Foundation of China [Grant No. 52192661], Heilongjiang Provincial Natural Science Foundation [Grant No. LH2022E070], and Fundamental Research Funds for the Central Universities [Grant No. HIT.NSRIF202334].

REFERENCES

[1] Xu, Y., Qian, W., Li, N., & Li, H. (2022). Typical advances of artificial intelligence in civil engineering. Advances in Structural Engineering, 25(16): 3405-3424.
[2] Xu, Y., Li, S., Zhang, D., Jin, Y., Zhang, F., Li, N., & Li, H. (2018). Identification framework for cracks on a steel structure surface by a restricted Boltzmann machines algorithm based on consumer-grade camera images. Structural Control and Health Monitoring, 25(2), e2075.
[3] Xu, Y., Wei, S., Bao, Y., & Li, H. (2019). Automatic seismic damage identification of reinforced concrete columns from images by a region-based deep convolutional neural network. Structural Control and Health Monitoring, 26(3), e2313.
[4] Xu, Y., Li, Y., Zheng, X., Zheng, X., & Zhang, Q. (2023). Computer-vision and machine-learning-based seismic damage assessment of reinforced concrete structures. Buildings, 13(5), 1258.
[5] Xu, Y., Bao, Y., Chen, J., Zuo, W., & Li, H. (2019). Surface fatigue crack identification in steel box girder of bridges by a deep fusion convolutional neural network based on consumer-grade camera images. Structural Health Monitoring, 18(3), 653-674.
[6] Xu, Y., Fan, Y., & Li, H. (2023). Lightweight semantic segmentation of complex structural damage recognition for actual bridges. Structural Health Monitoring, 22(5), 3250-3269.
[7] Zhao, J., Hu, F., Qiao, W., Zhai, W., Xu, Y., Bao, Y., & Li, H. (2022). A modified U-Net for crack segmentation by Self-Attention-Self-Adaption neuron and random elastic deformation. Smart Structures and Systems, 29(1), 1-16.
[8] Wang, Y., Jing, X., Xu, Y., Cui, L., Zhang, Q., & Li, H. (2023). Geometry-guided semantic segmentation for post-earthquake buildings using optical remote sensing images. Earthquake Engineering & Structural Dynamics, 52(11), 3392-3413.
[9] Wang, Y., Cui, L., Zhang, C., Chen, W., Xu, Y., & Zhang, Q. (2022). A two-stage seismic damage assessment method for small, dense, and imbalanced buildings in remote sensing images. Remote Sensing, 14(4), 1012.
[10] Cui, L., Jing, X., Wang, Y., Huan, Y., Xu, Y., & Zhang, Q. (2022). Improved swin transformer-based semantic segmentation of postearthquake dense buildings in urban areas using remote sensing images. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 16, 369-385.
[11] Wang, Y., Jing, X., Cui, L., Zhang, C., Xu, Y., Yuan, J., & Zhang, Q. (2023). Geometric consistency enhanced deep convolutional encoder-decoder for urban seismic damage assessment by UAV images. Engineering Structures, 286, 116132.
[12] Xu, Y., Qiao, W., Zhao, J., Zhang, Q., & Li, H. (2023). Vision-based multi-level synthetical evaluation of seismic damage for RC structural components: a multi-task learning approach. Earthquake Engineering and Engineering Vibration, 22(1), 69-85.
[13] Xu, Y., Bao, Y., Zhang, Y., & Li, H. (2021). Attribute-based structural damage identification by few-shot meta learning with inter-class knowledge transfer. Structural Health Monitoring, 20(4), 1494-1517.
[14] Xu, Y., Fan, Y., Bao, Y., & Li, H. (2023). Task-aware meta-learning paradigm for universal structural damage segmentation using limited images. Engineering Structures, 284, 115917.
[15] Zhong, J., Fan, Y., Zhao, X., Zhou, Q., & Xu, Y. (2024). Multi-type structural damage image segmentation via dual-stage optimization-based few-shot learning. Smart Cities, 7(4), 1888-1906.
[16] Xu, Y., Fan, Y., Bao, Y., & Li, H. (2024). Few-shot learning for structural health diagnosis of civil infrastructure. Advanced Engineering Informatics, 2024, 62, A, 102650.
[17] Fan, Y., Li, H., Bao, Y., Xu, Y. (2024). Cycle-consistency-constrained few-shot learning framework for universal multi-type structural damage segmentation. Structural Health Monitoring.
[18] Liu, H., Jia, C., Shi, F., Cheng, X., & Chen, S. (2025). SCSegamba: Lightweight Structure-Aware Vision Mamba for Crack Segmentation in Structures. CVPR 2025, arXiv:2503.01113.