Smart Structures and Systems, *Vol. 33*, *No. 1* (*2024*) *55-64*
https://doi.org/10.12989/sss.2024.33.1.055

Coating defect classification method for steel structures with vision–thermography imaging and zero-shot learning

Jun Lee [a], Kiyoung Kim [b], Hyeonjin Kim [a] and Hoon Sohn*

Department of Civil Engineering, Korean Advanced Institute for Science and Technology, 291 Daehak-ro, Yuseong-gu, Daejeon 34141, Republic of Korea

(*Received October 31, 2023, Revised December 6, 2023, Accepted December 10, 2023*)

Abstract. This paper proposes a fusion imaging-based coating-defect classification method for steel structures that uses zero-shot learning. In the proposed method, a halogen lamp generates heat energy on the coating surface of a steel structure, and the resulting heat responses are measured by an infrared (IR) camera, while photos of the coating surface are captured by a charge-coupled device (CCD) camera. The measured heat responses and visual images are then analyzed using zero-shot learning to classify the coating defects, and the estimated coating defects are visualized throughout the inspection surface of the steel structure. In contrast to older approaches to coating-defect classification that relied on visual inspection and were limited to surface defects, and older artificial neural network (ANN)-based methods that required large amounts of data for training and validation, the proposed method accurately classifies both internal and external defects and can classify coating defects for unobserved classes that are not included in the training. Additionally, the proposed model easily learns about additional classifying conditions, making it simple to add classes for problems of interest and field application. Based on the results of validation via field testing, the defect-type classification performance is improved 22.7% of accuracy by fusing visual and thermal imaging compared to using only a visual dataset. Furthermore, the classification accuracy of the proposed method on a test dataset with only trained classes is validated to be 100%. With word-embedding vectors for the labels of untrained classes, the classification accuracy of the proposed method is 86.4%.

Keywords: active thermography; defect inspection; non-destructive test; steel structure; zero-shot learning

1. Introduction

Coatings are widely used to protect steel structures from various forms of deterioration such as corrosion, abrasion, and cracking (Standard 1994, Shrestha and Kim 2018). Generally, coatings are applied onto a surface in a certain thickness to ensure protective performance on the structure. For this purpose, polymeric materials have mainly been used for their high applicability. However, owing to their thinness and softness properties, which are beneficial to their workability, such coatings easily become defective from external shocks and gradually lose their protective performance. Additionally, the coating layer undergoes a reduction in thickness and quality over time, resulting in several defects, such as corrosion, delamination, chalking, and checking (Ochiai *et al*. 2007, A.A.o.S.a.H.T. Officials 2019, Center 2021, Boller *et al*. 2015). Hence, the timely and accurate detection of coating defects is crucial, and visual inspection has become a mandatory practice in the maintenance of steel bridges, as stipulated by bridge authorities such as the Ministry of Land, Infrastructure, and Transportation in the Republic of Korea and the Federal Highway Administration in the United States (Jeong *et al*. 2018, Puspitasari and Harahap 2023).

Visual inspection is a prominent method in quality and cleanliness assessment procedures and is generally employed in on-site coating-defect inspection. It is strongly dependent on the level of experience and cognitive concentration of the inspector, making it a subjective process with high potential for human error. To address this limitation, visual-camera-based technologies have been applied to defect detection (Hoult *et al*. 2010, La *et al*. 2019, Jiang *et al*. 2023). Unmanned aerial vehicles (UAVs) and automated robots have recently been adopted to comprehensively inspect the structural members of steel bridges by capturing high-resolution visual images (Myung *et al*. 2014, Lee *et al*. 2021). Subsequently, these images are organized to automate the diagnostic process for efficient and accurate assessment. There have also been various studies on the coating-defect inspection of ferrous protection coatings in automotive and machinery industries. Common methods include the use of ultrasound (Li and Zhang 2008, Ulbrich 2022), microwave (Deshmukh *et al*. 2011, Mazzinghi *et al*. 2019), and X-ray in inspections (Margret *et al*. 2018). However, these techniques suffer from drawbacks in that they require large equipment and that their accuracies are vulnerable to external environmental conditions.

Recently, machine-learning techniques based on visual images from charge-coupled device (CCD) cameras have

*Corresponding author, Ph.D., Professor,
E-mail: hoonsohn@kaist.ac.kr
[a] Ph.D. Student
[b] Ph.D.

ISSN: 1738-1584 (Print), 1738-1991 (Online)

been employed to enhance the accuracy and efficiency of detecting cracks in coatings and corrosion in steel materials. For example, Jahanshahi and Masri (2013) conducted an assessment of the effects of employing various color spaces, color channels, and image patch dimensions in a color wavelet-based texture analysis algorithm aimed at corrosion detection and demonstrated that CbCr leads to optimal performance in corrosion detection in their texture analysis method. Ali and Cha (2019) proposed an approach to subsurface damage detection in steel-truss bridges using deep learning and infrared thermography. Their method effectively identified corrosion and detected debonding between the coating paint and the steel surface. Jin Lim *et al*. (2021) proposed a corrosion inspection method for steel structures that uses both visual and thermographic images. Thermal inspection can be used to effectively visualize, through heat transfer, the corrosion occurring inside and outside a steel structure. As such, thermal inspection has been integrated with visual-image-based technologies, which are incapable of seeing internal corrosion, to detect both external and internal defects. However, these algorithms have been trained and tested primarily on individual coating defects, which may pose an unknown limitation in detecting new classes of defects that have not been included in the training.

In this study, a coating-defect classification method based on zero-shot learning is proposed. The proposed method uses a fusion-imaging inspection system that combines halogen-lamp-based active thermography and vision-based inspection. The active thermography system consists of a halogen lamp that transmits heat over a large area of the structure, and an infrared (IR) camera that captures a series of thermal images. The proposed method employs zero-shot learning, which classifies untrained classes by embedding matrices from coating-defect-related documents. By utilizing zero-shot learning, the proposed method obtains a similarity metric between image features and class label features from documents related to coating defects. The proposed method (1) detects coating defects on surfaces and subsurface at high accuracies, (2) classifies the coating defects of larger areas more accurately than do older methods, and (3) classifies new coating defects that have not been inspected and untrained defects.

The remainder of this paper is organized as follows: An overview of the proposed fusion-imaging system and data is provided in Section 2. The zero-shot learning used in the coating defect classification method is described in detail in Section 3. Experimental validation is provided in Section 4. Finally, the concluding remarks are presented in Section 5.

2. Fusion imaging via thermography and visual inspection

Although thermography alone can be used to detect deterioration to a sufficient degree, it is challenging to differentiate the type of deterioration on the surface of a structure using only thermal imaging. To address this problem, visual images containing surface characteristics can be combined with thermography. This section introduces a detailed procedure and theoretical background of the fusion of these two imaging types.

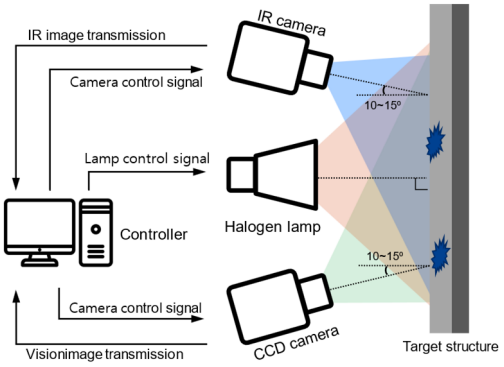

Fig. 1 Schematic of proposed fusion-imaging system combining visual- and thermal-imaging data for coating-defect classification

2.1 Configuration of fusion-imaging system

Fig. 1 shows the fusion-imaging system adopted for the proposed method and its working procedure, which was developed by Jin Lim *et al*. (2021). The system comprises four devices for heating, sensing, and control. A halogen lamp, which generates heat energy on the target surface, is placed at the center of the target structure. An IR camera captures a series of IR images that visualize the resulting heat responses of the target surface within its field of view (FOV). To avoid measuring the halogen light reflected from the target structure, the IR camera is mounted at an angle of 10–15° from the normal line of the target structure. A CCD camera captures photos of the surface at the target location. The controller operates by turning the halogen lamp on and off, transmitting control signals to the IR and CCD cameras, and receiving real-time IR images from the IR camera and visual images from the CCD camera to detect and classify coating defects.

In the proposed halogen thermography system, the halogen lamp is repeatedly turned on and off at specific periods with respect to an ON/OFF-type excitation signal. The monitoring time is divided into heating and cooling phases; for the proposed method, the durations of the two phases are set to be identical. Because the temperature change in the target structure should be discerned by the IR camera during each measurement, the monitoring time can be adjusted such that the surface temperature of the target structure increases by at least 3 °C in the test environment.

The controller in the proposed fusion-imaging system transmits control signals to the halogen lamp and two cameras to collect thermal- and visual-imaging data. After transmitting the operation signal to the CCD camera, the controller determines the monitoring time based on the test environment conditions, and transmits the lamp operation signal simultaneously to the halogen lamp and IR camera. The halogen lamp is then turned on and off in response to the signal from the controller to radiate heat onto the surface of the target structure, while the IR camera captures the thermal changes in the target structure in succession. The thermal images captured by the IR camera and the

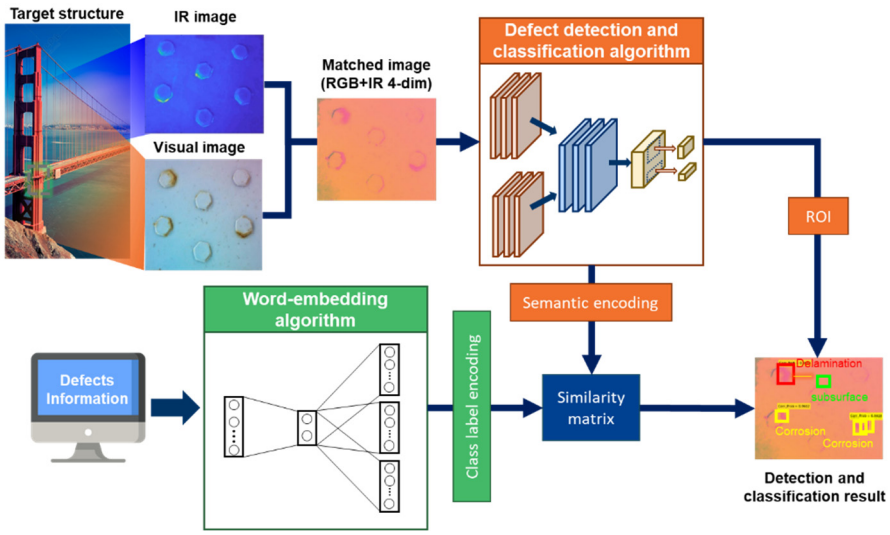

Fig. 2 Overview of coating-defect classification method for unobserved class dataset

visual images captured by the CCD camera are transferred to the controller and preprocessed to generate input data for the coating-defect classification.

2.2 Fusion of thermography and visual imaging

When a halogen lamp applies heat energy to the surface of a target structure, photothermal effect-based thermal waves are generated on the surface and propagate inside the structure. The presence of corrosion modifies the thermal conductivity, reflectivity, and emissivity of a material, generating spatial and temporal fluctuations in the propagation of thermal waves around the corrosion. In addition, subsurface corrosion, which occurs frequently at the interface between the coating layer and the steel structure, disrupts the transmission of heat waves. Lock-in amplitude thermal imaging, referred to hereafter in this paper as 1D lock-in amplitude thermal imaging, is a method for visualizing temporal and spatial disruptions. When heat energy is applied to a structure for $t/2$ seconds and thermal images are recorded over t seconds, the 1D lock-in amplitude thermal image $A(x,y)$ of a point with surface-plane Cartesian coordinates (x,y) is obtained as follows

$$A(x,y) = \sqrt{\left\{R(x,y,0) - R\left(x,y,\frac{t}{2}\right)\right\}^2 + \left\{R\left(x,y,\frac{t}{2}\right) - R(x,y,t)\right\}^2}, \quad (1)$$

where $R(x,y,t)$ is the temperature of the raw thermal image recorded by the IR camera, and $A(x,y,t)$ is the 1D lock-in amplitude thermal image at (x,y) at time t.

In previous studies on steel-structure examination using visual sensors, color spaces such as gray scale or YCbCr have been used to represent corrosion, which is an important defect in steel structures (Liao and Lee 2016, Khayatazad *et al.* 2020). In accordance with these color spaces, the texture of the corrosion is distinctly rendered to be able to perform target-surface inspection at high precisions. However, corrections for accuracy have to be made by adjusting the associated parameters for each target structure and inspection environment. In this study, the well-known three-dimensional color map, which consists of red, blue, and green, is used for generalization because the proposed method for classifying coating defects is intended to be able to detect defects that occur in steel structures, other than corrosion.

3. Coating-defect classification method for unobserved-class dataset

Fig. 2 depicts the coating-defect classification method for unobserved-class datasets. The proposed method consists of two algorithms: a defect detection/classification algorithm and a word-embedding algorithm, which are represented in the diagram in orange and green, respectively. In the visual-feature extraction algorithm, a four-dimensional image dataset containing visual- and thermal-imaging data is used to extract visual-feature vectors and identify the region of interest (ROI). In the word-embedding algorithm, the coating-defect data are then converted into vectors, and label data are generated in the embedding space. In this section, the YOLOv5-based visual-feature extraction algorithm and word-embedding algorithm based on Word2vec are explored in detail. The similarity between a visual-feature vector and word vector is determined by transforming the former to the same size as that of the word-embedding vector and comparing it with the word vector. The vector with the highest similarity is identified as the defect class, and the defect location and size are visualized by outputting the ROI results.

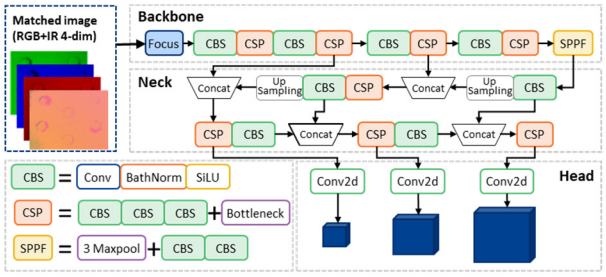

Fig. 3 Overview of YOLOv5s-based visual-feature extraction and region of interest

3.1 Visual/thermal image-feature extraction algorithm

You Only Look Once (YOLO) is a method that detects objects in real time within a single stage (Redmon *et al.* 2016). Through the use of a convolutional neural network (CNN) architecture on an image, it is possible to determine the positions and types of objects in the image, thereby accelerating recognition. YOLO divides an image into grids, where each grid detects the objects contained within itself. This method can be used for real-time object detection based on data streams and requires only minimal computational resources. The coating-defect classification method utilizes the fifth version of YOLO, which is still under development, called YOLOv5. YOLOv5 is distinguished from its previous versions by the depth of the depth-multiplier control model and different input image-data sizes. In this study, YOLOv5s, the most basic version with the lightest file, was used to achieve the fastest recognition speed. Fig. 3 shows the structure of YOLOv5s.

Like any other single-stage object detector, YOLOv5s has three important components: (1) backbone, (2) neck, and (3) head. The backbone is used primarily to extract significant features from images. The specific structures of each component is shown in Fig. 3. In YOLOv5s, the backbone and neck are designed as cross-stage partial networks (CSP) to extract informative features from an input image. Using CSP in deeper networks significantly reduces processing time. The primary function of the neck is to generate feature pyramids, which facilitate the generalization of models for object scaling. The feature pyramid consists of three feature vectors of various sizes and scales. When multiple sizes and scales of feature vectors are used, it becomes easier to detect the same object. The feature vectors from the neck are used for semantic encoding and are transformed to be of the same size as that of the word-embedding vector. Finally, the model head is employed primarily in the final phase of detection. It generates the final output vectors with class probabilities, objectless scores, and bounding boxes.

3.2 Coating-defect-type embedding

Word embedding involves the conversion of a single word into a numerical representation. Each word is mapped to a vector, which is then learned in a manner similar to that used by a neural network. The vectors attempt to capture various aspects of a word in relation to the entire text, including its semantic relationship, definition, and context. By utilizing these numerical representations, one can identify similarities and differences between words, among other aspects. Text in its raw form cannot be processed by a machine; therefore, the text has to be converted into an embedding to enable users to feed it into traditional machine-learning models.

According to past studies (A.A.o.S.a.H.T. Officials 2019, I.a.T.o.S.K. Ministry of Land 2019), the three most common defects in coating materials used to protect steel structures are corrosion, delamination, and subsurface problems. Based on the maintenance documents mentioned above and web crawling results from Google search, 75 words were chosen in order of high relevance for each coating defect in this study. Omitting terms with minimal relevance, the side information list had a total of 150 words, as shown in Fig. 4.

The word-embedding algorithm is based on the skip-gram model from the Word2vec method. Word2Vec is effective because of its capacity to group together vectors of similar words. Word2Vec can accurately estimate the meaning of a word based on its occurrences in text, given a sufficiently large dataset. Fig. 4 shows a schematic representation of the neural network of the skip-artificial gram. In the input layer, a one-hot vector of the central

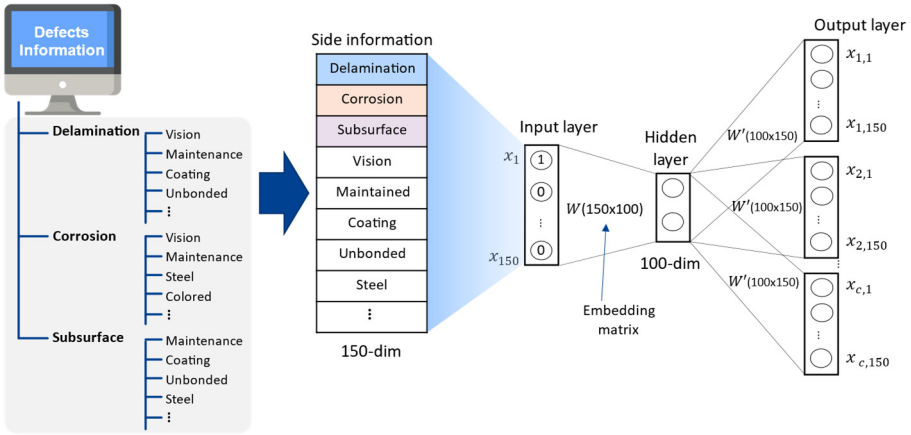

Fig. 4 Schematic of word-embedding methods for coating-defect types

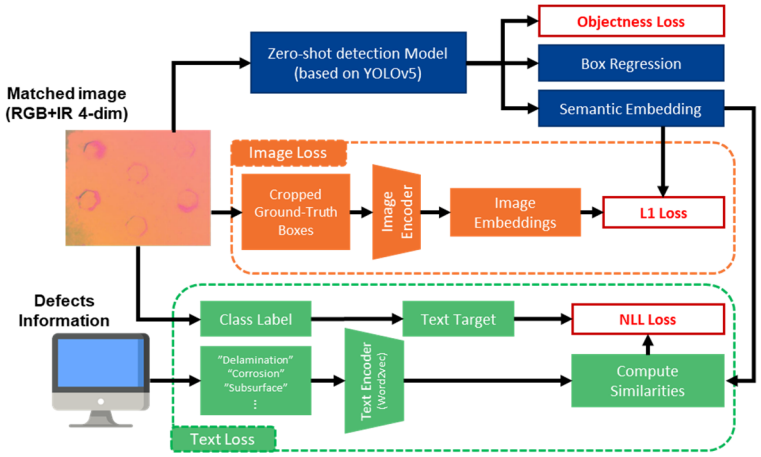

Fig. 5 Loss function of coating-defect classification method based on YOLOv5 and Word2vec

word is introduced. When the input passes through the hidden layer, the output layer generates a vector that predicts the surrounding words. As shown in Fig. 4, the skip-gram is not a deep artificial neural network model with many hidden layers, but rather a shallow neural network model with only one hidden layer.

The goal of skip-gram is to predict surrounding words for the central word by learning the embedding matrix W between the input layer and hidden layer, and the weight matrix W' between the hidden layer and output layer. For training, the loss function is calculated as follows

$$x_\alpha = \text{softmax}(W'Wx) \ \ (\alpha = 1, 2, \ldots, c). \tag{2}$$

3.3 Full dual loss function

In the visual-feature extraction method, the final image loss function is calculated using a simple L1 loss function. Let the semantic embedding outputs for ground-truth matched anchors be denoted by M_g, as shown in Fig. 5 and let the matching target vector embedding be denoted by I_g. The fundamental image loss function is expressed as follows

$$L_i = \text{mean}\left(\left|(M_g - I_g)\right|\right). \tag{3}$$

In the word-embedding method, the process of distilling text embeddings involves the alignment of model semantic outputs with the target text embeddings. To compute the loss for the text embeddings, the seen text embeddings are first generated by feeding every seen class. Then, a similarity matrix is generated using a cosine similarity computation relating semantic embeddings M_t and seen text embedding T. Additionally, a softmax is applied with temperature τ to these similarities. The full computation to generate the similarity matrix S_z can be expressed as follows

$$S_z = \text{softmax}\left(\frac{M_t^T T}{\|M_t\|\|T\|} e^\tau\right). \tag{4}$$

To compute the overall text loss, we first collect the ground-truth box label classes that correspond to each positively matched anchor. Then, we produce the output layer of the word embedding, which are one-hot encoded labels, y, whose shape is the same as that of the generated similarity matrix S_z. The final text loss is computed using a function called the negative log-likelihood, which links S_z and y, and can be expressed as follows

$$L_t = L_{NLL}(\log(S_z), y). \tag{5}$$

Combining both text and image distillation losses into a single function with weighting values W_t and W_i produces a full-loss function, as follows

$$Loss = W_t L_t + W_i L_i. \tag{6}$$

4. *In situ* bridge testing

4.1 Test configuration

The performance of the proposed method was validated on the Gwangan Bridge, Deungsun Bridge, and First Jindo Grand Bridge in South Korea. The Gwangan Bridge, an offshore bridge completed in 2002, is composed of a steel-truss bridge (900 m) and a suspension bridge (720 m), each with a width of 25 m. The First Jindo Grand Bridge is a three-span steel-box girder cable-stayed bridge with a long main span (344 m) and two side spans (70 m), each with a width of 11.7 m. The bridge was constructed on the ocean and completed in 1984. The Deungsun Bridge is composed of steel box girders and prestressed concrete box girders. The entire length of the bridge is 2,000 m, and the span length is 11 m. The bridge was constructed near a river and was completed in 1984.

Because the Gwangan Bridge and First Jindo Grand Bridge are offshore structures, they are usually exposed to a strong sea breeze, and as a result, their coatings deteriorate quickly and easily develop flaws, particularly corrosion around bolts and welds. The steel girders of the Deungsun Bridge are also susceptible to defects owing to the intense vortices caused by topographical influences. Because of these environmental threats, it is quite probable that, for these three bridges, damage has gradually accumulated not only on the coatings applied to the structures but also in the

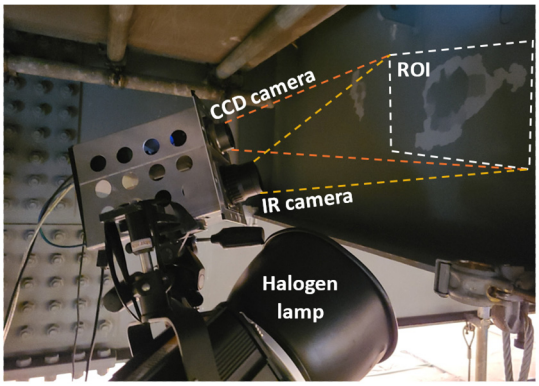

Fig. 6 Hardware configuration of fusion-imaging system for in situ bridge inspection

structures themselves. Therefore, an effective coating-defect classification diagnostic method is essential to protect the structural integrities of buildings and ensure effective defect maintenance.

Fig. 6 shows the fusion-imaging system used for *in situ* bridge inspection, consisting of a combined visual–IR camera (T650SC, FLIR), a halogen lamp (H25 S, Hedler), and a laptop (XPS159570, Dell). The combined visual–IR camera acquires visual and thermal images with 95 × 71 cm^2 and 33 × 44 cm^2 FOVs, respectively, and 2592 × 1944-pixel and 640 × 480-pixel resolutions, respectively. The thermal resolution and spectral range of the thermal image are 0.02 K and long-wave IR (7.5–13 mm), respectively. The fusion-imaging system was placed 1 m from the target surface, based on a compromise between the inspection speed and spatial resolution of the acquired images. To ensure uniform heat excitation, the halogen lamp was set 90° to the target surface. To minimize the reflected heat of the halogen lamp, the IR and CCD cameras were set at 20° from the target surface. Before the halogen lamp introduced heat energy onto the target surface, the CCD camera captured the visual image. The halogen lamp then applied heat for 10 s, and thermal images were recorded for 20 s (10 s for the heating phase and additional 10 s for the cooling phase) at a 30 Hz frame rate. These parameter values were selected by considering the halogen lamp power, external temperature, and thermal properties of the coating material. The 1D amplitude thermal image was obtained using Eq. (1). The four-dimensional input, which is a fusion of 3D visual images and 1D amplitude thermal images, was then obtained.

Table 1 Number of fusion-imaging datasets used for data augmentation

Defect type	Original data	Data augmentations	Total defects
Delamination	5	40	56
Corrosion	15	120	256
Subsurface	5	40	64

Detailed information regarding the images acquired from the bridges cannot be published in this article because of confidentiality agreements with the bridge authority.

4.2 Training of coating-defect classification method

The proposed method was coded in Python using the PyTorch framework (Paszke *et al*. 2019). The main network of the proposed method was deployed on a workstation with the following technical specifications: (1) CPU: Intel Core i7-9700, (2) GPU: Nvidia Force 2060 Super 8 GB, and (3) RAM: 32 GB.

A transfer-learning method was utilized to train the proposed method. The visual-feature extraction model was pretrained using the VGGNet dataset, which consists of 1.2M images with assigned categories, whereas the word-embedding model was pretrained using 71,380 words from the Google data center.

Tabe 1 lists the fault types and amounts of data for each detection type for the *in situ* bridge inspection. One problem was that the number of defects required for training was insufficient. Because deep learning is prone to overfitting, and the validation accuracy is diminished under a limited amount of training data, the proposed method expands the dataset via data augmentation.

Fig. 7 shows seven types of data augmentation approaches. To train the word embedding of the defect-class labels, only related words were selected using coating-defect-related documents. There were four classes of interest: normal condition, delamination, corrosion, and subsurface. However, because the same learning was required for the other surrounding words, the input vector size was set to 150, the hidden layer to 100 dimensions, and the output vector size to 4 × 150 dimensions with four peripheral words.

The subsurface class was set to be an unobserved class. Thus, to test the performance under the unobserved-class dataset, image data from the subsurface class were not used

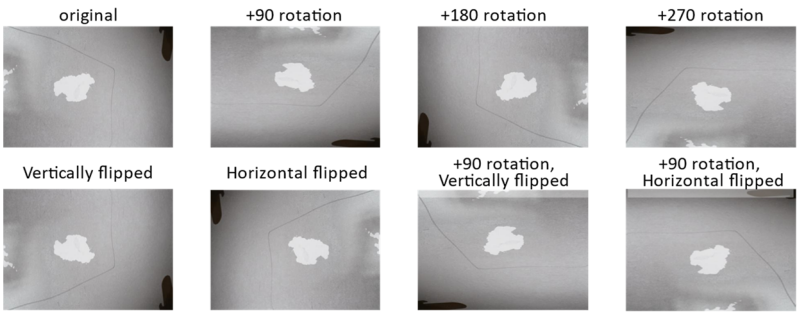

Fig. 7 Data augmentation of fusion-imaging dataset

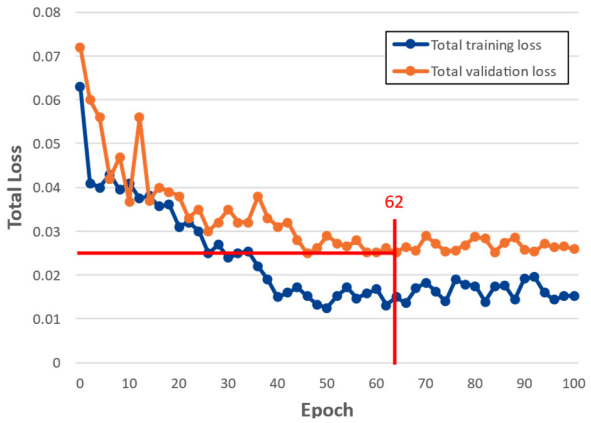

Fig. 8 Training and validation losses of proposed method

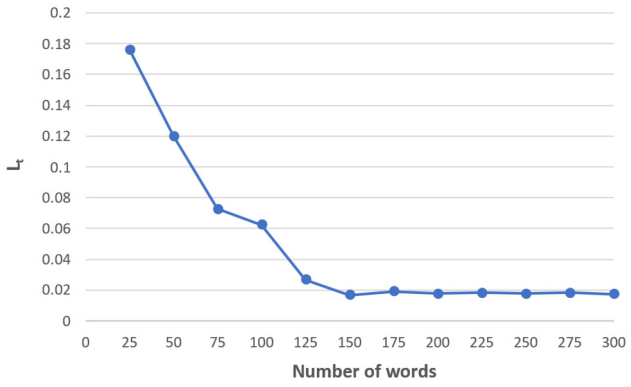

Fig. 9 L_t loss corresponding to different numbers of words

for training. Of the 200 fusion-imaging datasets, 100 and 28 were used for training and validation, respectively. The remaining 72 fusion-imaging datasets, which included the subsurface class, were used for testing.

The proposed methods were trained using an Adam optimizer with a total of 100 epochs at 50 iterations per epoch, a starting learning rate of 0.0001, weight decay of 0.001, momentum of 0.9, and batch size of 4. Fig. 8 shows the training process for the proposed method. After epoch 62, the validation loss no longer decreased and began to oscillate. Therefore, training was stopped at epoch 62 to prevent overfitting.

The harmonic mean (HM) is extensively used to evaluate the performance of zero-shot learning-based object classification. The HM value was calculated based on the accuracies on the observed and non-observed data. In the zero-shot learning model, training is performed on the seen classes, inevitably resulting in a bias toward those specific classes. To account for a decrease in the HM score when the accuracies for unseen classes decrease to below those for the seen ones, adjustments were made, as in Eq. (7)

$$\mathrm{HM} = 2 * \frac{Acc_s * Acc_u}{Acc_s + Acc_u}, \tag{7}$$

where Acc_s and Acc_u denote the accuracies for the seen and unseen classes, respectively.

4.3 Test results

Fig. 9 shows the resulting L_t losses corresponding to different numbers of words. Semantic embedding, which is essential for L_t analysis, was conducted using the optimized results obtained from YOLOv5, which were used in the development model. The experiment was conducted 50 times, and the results were averaged. The numbers of words were varied from 25 to 300 at intervals of 25. The experimental results indicated a clear tendency for L_t to decrease as the number of words is increased. This observation implies an increasing similarity among the embedding results, indicating the potential for higher accuracy in classifying unobserved data. However, the L_t value converged and stabilized after 125 words, which implies that exceeding this threshold would result in unnecessary parameters in the word-embedding process. Therefore, in this study, 150 words were used to achieve the lowest loss.

The defect-classification performance was evaluated in terms of the accuracy for different input data types, and the results are shown in Table 2. To validate the performance of the proposed method, four types of input data were used. 3D RGB input data are used in common vision-based defect-classification methods, whereas 1D IR input data are used as amplitudes for thermography-based defect classification methods. In Lim's method (Jin Lim *et al.* 2021), the 3D CbCr/IR data are combined with 2D vision data, which are converted by the CbCr method to form RGB and 1D IR data. Compared to the use of fusion data, the use of only one of either visual- or thermal-imaging data resulted in lower average accuracies. In the case of single-data-type usage, it was difficult to determine whether the coating defects were inside or on the surface of the coating. By contrast, when combined visual and infrared image data were used as input, the coating defects were accurately identified. However, in terms of Acc_u, CbCr was approximately 5% less accurate than the proposed method because of the information lost during the conversion from RGB. Of the input data types that were compared, the 4D RGB/IR input data type, used by the proposed method, resulted in a high-level HM of 92.6%.

Subsequently, the defect-classification performance on the observed- and unobserved-class datasets was evaluated using accuracy as a metric, and the results are shown in Table 3. The accuracy of the proposed method was compared with those of state-of-the-art visual-feature extraction models, i.e., Mask R-CNN (He *et al.* 2017), DenseNet-201 (Zhu and Newsam 2018), ResNet-50 (He *et*

Table 2 Results for input data types, after 150 independent training sessions

Input data type	Average accuracy		HM
	Acc_s	Acc_u	
3D RGB	87.2	63.7	73.6
1D IR	82.5	75.2	78.7
3D CbCr/IR	100	81.3	89.7
4D RGB/IR	100	86.4	92.7

al. 2016), EfficientNet (Koonce 2021), and YOLOv5 (the proposed model). For a comprehensive comparison of image-feature extraction models, we employed the word-embedding model used in our proposed approach. The input data for this evaluation were 4D RGB/IR data.

As for the accuracy on the observed data, the values for each model did not show a significant difference; the highest and lowest accuracies of the models had a difference of only 4%. However, when the accuracy on the non-observed newly introduced class was evaluated, there was a noticeable difference in accuracy of approximately 12%. In this study, the object detection involved only four classes, and the limited number of classes is believed to have contributed to the relatively small difference. By contrast, the accuracy on the non-observed data was attributed to the implemented feature vectors, particularly in relation to the proposed word-embedding vectors and their similarity. These findings demonstrate the performance of the YOLOv5 model, making it the most suitable choice for this study on four-dimensional data.

Three samples of defect-classification results from the proposed method are shown in Fig. 10. Accurate detections were achieved on the delamination and corrosion datasets. Additionally, the ROI results showed that the types and sizes of defects on the planar surface were correctly detected. In the corrosion results, the detection region was smaller in the thermal image than in the visual image. Nonetheless, the detection results revealed that the corrosion-related ROI was accurately recognized. Thus, the defect-classification accuracy using both thermal and visual images was higher than that using individual data. The results for the unknown-class dataset are shown in the right column of Fig. 10. The ROI location and size results for the subsurface class, highlighted by a red box, were determined with high precision. However, in the case of minor areas, there was no dataset learning; thus, observed-class errors occurred.

The accuracies of the results for non-observed data are compared in Table 4. The accuracy was at its lowest when the subsurface class was the non-observed class and at its highest when corrosion was the non-observed class. As shown in Fig. 10, an error that occurred in the subsurface was misclassified as delamination because subsurface defects and delamination share similar characteristics, with the feature vectors exhibiting relatively close distances. Hence, errors occur when the features cannot be clearly distinguished owing to their small size. However, this size-related error problem is expected to be easily addressed by increasing the resolutions of images for both visual and thermal imaging when the feature sizes are small, allowing for better feature determination.

Table 3 Results of different image-feature extraction models after 150 independent training session (same model of word embedding)

Input data type	Average accuracy		HM
	Acc_s	Acc_u	
Mask R-CNN	96.2	73.7	83.5
DenseNet-201	98.2	71.2	82.5
ResNet-50	97.5	83.1	89.7
EfficientNet	100	83.3	90.9
YOLOv5 (our model)	100	86.4	92.7

Table 4 Results for non-observed class, after training with 50 data points

Classes		Acc_u	HM
Observed	Non-observed		
Corrosion subsurface	Delamination	85.5	92.2
Delamination subsurface	Corrosion	90.7	95.1
Delamination corrosion	Subsurface	83.1	90.8

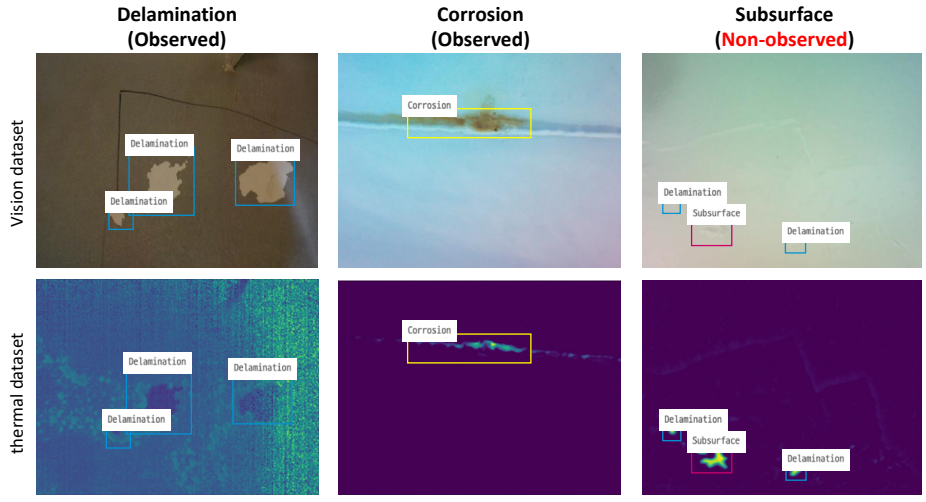

Fig. 10 Defect-classification results including non-observed class dataset (subsurface class is non-observed)

5. Conclusions

In this study, a coating-defect classification method for unobserved-class datasets based on zero-shot learning was proposed. The proposed method utilizes visual and thermal images obtained *in situ* from bridges, labeled to form a training dataset. Typical coating-defect labels were obtained from maintenance guidelines. Based on the validation results, the defect-type classification performance was improved by fusing visual and thermal imaging compared with using only a visual dataset. Moreover, the classification accuracy of the proposed method on a test dataset with only observed classes was validated to be 100%. With word-embedding vectors for the labels of unobserved classes, the classification accuracy of the proposed method was 86.4%. As more training data become available, the precision and robustness of the proposed method should increase.

Chalking and checking inspections for coating defects should also be performed. The thermal properties of these defects differ from those of the defects used in this study. Based on this observation, improvements in machine-learning-based coating-defect classification techniques are being developed using visual and thermal fusion data. In subsequent studies, new findings will be revealed. We are also minimizing the size, weight, and power consumption of the fusion-imaging system for bridge inspection and developing portable coating-defect inspection devices based on the proposed method. The comparatively small sizes and light weights of these devices will enable simultaneous data acquisition and signal processing at bridge maintenance sites.

Acknowledgments

This work was supported by a National Research Foundation of Korea (NRF) grant funded by the Korean government (MSIT) [Grant Number 2019R1A3B3067987].

References

A.A.o.S.a.H.T. Officials (Ed.) (2019), Standard Practice for Evaluation of Coating Systems for Structural Steel.

Ali, R. and Cha, Y.J. (2019), "Subsurface damage detection of a steel bridge using deep learning and uncooled micro-bolometer", *Constr. Build. Mater.*, **226**, 376-387. https://doi.org/10.1016/j.conbuildmat.2019.07.293

Boller, C., Starke, P., Dobmann, G., Kuo, C.M. and Kuo, C.H. (2015), "Approaching the assessment of ageing bridge infrastructure", *Smart Struct. Syst., Int. J.*, **15**(3), 593-608. https://doi.org/10.12989/sss.2015.15.3.593

Center, N.L.I. (2021), Detailed guidelines for safety and maintenance of facilities.

Deshmukh, S., Xu, X., Mohammad, I. and Huang, H. (2011), "Antenna sensor skin for fatigue crack detection and monitoring", *Smart Struct. Syst., Int. J.*, **8**(1), 93-105. https://doi.org/10.12989/sss.2011.8.1.093

He, K., Zhang, X., Ren, S. and Sun, J. (2016), "Deep residual learning for image recognition", *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pp. 770-778. https://doi.org/10.1109/CVPR.2016.90

He, K., Gkioxari, G., Dollár, P. and Girshick, R. (2017), Mask R-CNN, pp. 2961-2969.

Hoult, N.A., Fidler, P.R., Hill, P.G. and Middleton, C.R. (2010), "Wireless structural health monitoring of bridges: Present and future", *Smart Struct. Syst., Int. J.*, **6**(3), 277-290. https://doi.org/10.12989/sss.2010.6.3.277

Jahanshahi, M.R. and Masri, S.F. (2013) "Effect of color space, color channels, and sub-image block size on the performance of wavelet-based texture analysis algorithms: An application to corrosion detection on steel structures", In: *Computing in Civil Engineering - Proceedings of the 2013 ASCE International Workshop on Computing in Civil Engineering*, pp. 685-692. https://doi.org/10.1061/9780784413029.086

Jeong, Y., Kim, W., Lee, I. and Lee, J. (2018), "Bridge inspection practices and bridge management programs in China, Japan, Korea, and U.S.", *J. Struct. Integr. Maint.*, **3**(2), 126-135. https://doi.org/10.1080/24705314.2018.1461548

Jiang, S., Wu, Y. and Zhang, J. (2023), "Bridge coating inspection based on two-stage automatic method and collision-tolerant unmanned aerial system", *Automat. Constr.*, **146**, 104685. https://doi.org/10.1016/j.autcon.2022.104685

Jin Lim, H., Hwang, S., Kim, H. and Sohn, H. (2021), "Steel bridge corrosion inspection with combined vision and thermographic images", *Struct. Health Monitor.*, **20**(6), 3424-3435. https://doi.org/10.1177/1475921721989407

Khayatazad, M., De Pue, L. and De Waele, W. (2020), "Detection of corrosion on steel structures using automated image processing", *Develop. Built Environ.*, **3**, p. 100022. https://doi.org/10.1016/j.dibe.2020.100022

Koonce, B. (2021), "EfficientNet", In: *Convolutional Neural Networks with Swift for Tensorflow*, Berkeley, CA, USA, pp. 109-123. https://doi.org/10.1007/978-1-4842-6168-2_10

Kreislova, K. and Geiplova, H. (2012), "Evaluation of corrosion protection of steel bridges", *Procedia Eng.*, **40**, 229-234. https://doi.org/10.1016/j.proeng.2012.07.085

La, H.M., Dinh, T.H., Pham, N.H., Ha, Q.P. and Pham, A.Q. (2019), "Automated robotic monitoring and inspection of steel structures and bridges", *Robotica*, **37**(5), 947-967. https://doi.org/10.1017/S0263574717000601

Lee, J.H., Yoon, S., Kim, B., Gwon, G.H., Kim, I.H. and Jung, H.J. (2021), "A new image-quality evaluating and enhancing methodology for bridge inspection using an unmanned aerial vehicle", *Smart Struct. Syst., Int. J.*, **27**(2), 209-226. https://doi.org/10.12989/sss.2021.27.2.209

Li, X. and Zhang, Y. (2008), "Feasibility study of wide-band low-profile ultrasonic sensor with flexible piezoelectric paint", *Smart Struct. Syst., Int. J.*, 565-582. https://doi.org/10.12989/sss.2008.4.5.565

Liao, K.W. and Lee, Y.T. (2016), "Detection of rust defects on steel bridge coatings via digital image recognition", *Automat. Constr.*, **71**(Part 2), 294-306. https://doi.org/10.1016/j.autcon.2016.08.008

Margret, M., Menaka, M., Subramanian, V., Baskaran, R. and Venkatraman, B. (2018), "Non-destructive inspection of hidden corrosion through Compton backscattering technique", *Radiat. Phys. Chem.*, **152**, 158-164. https://doi.org/10.1016/j.radphyschem.2018.07.015

Mazzinghi, A., Freni, A. and Capineri, L. (2019), "microwave non-destructive testing method for controlling polymeric coating of metal layers in industrial products', *NDT E Int.*, **102**, 207-217. https://doi.org/10.1016/j.ndteint.2018.12.003

Ministry of Land (2019), *Guidelines for the implementation of safety maintenance of infrastructure*. https://www.mlit.go.jp/road/road_e/s3_maintenance.html

Myung, H., Wang, Y., Kang, S.C. and Chen, X. (2014), "Survey on robotics and automation technologies for civil infrastructure".

http://ir.canterbury.ac.nz/handle/10092/10781
Ochiai, S., Iwamoto, S., Nakamura, T. and Okuda, H. (2007) "Crack spacing distribution in coating layer of galvannealed steel under applied tensile strain", *ISIJ Int*., **47**(3), 458-465. https://doi.org/10.2355/isijinternational.47.458
Paszke, A., Gross, S., Massa, F., Lerer, A., Bradbury, J., Chanan, G., Killeen, T., Lin, Z., Gimelshein, N., Antiga, L. and Desmaison, A. (2019), *PyTorch*: *An imperative style*, *high-performance deep learning library*, *Advances in Neural Information Processing Systems*. http://papers.nips.cc/paper/9015-pytorch-an-imperative-stylehigh-
Puspitasari, S.D. and Harahap, S. (2023), "Bridge inspection implementations and maintenance planning-A comparative analysis of a few distinctive countries", In: *AIP Conference Proceedings*, Vol. 2482, p. 50007. https://doi.org/10.1063/5.0110943
Redmon, J., Divvala, S., Girshick, R. and Farhadi, A. (2016), "You only look once: Unified, real-time object detection", *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pp. 779-788. https://doi.org/10.1109/CVPR.2016.91
Shrestha, R. and Kim, W. (2018), "Evaluation of coating thickness by thermal wave imaging: A comparative study of pulsed and lock-in infrared thermography – Part II: Experimental investigation", *Infrared Phys*. *Technol*., **92**, 24-29. https://doi.org/10.1016/j.infrared.2018.05.001
Standard (1994), "Surface preparation and protective coating", *M-Cr-501*, Rev. 1(December), pp. 20–31. https://www.standard.no/pagefiles/1167/m-501.pdf
Ulbrich, D. (2022), "Monitoring the boundary of an adhesive coating to a steel substrate with an ultrasonic Rayleigh wave", *Open Eng*., **12**(1), 933-945. https://doi.org/10.1515/eng-2022-0383
Zhu, Y. and Newsam, S. (2018), "DenseNet for dense flow", In: *Proceedings - International Conference on Image Processing*, *ICIP*, IEEE Computer Society, pp. 790-794. https://doi.org/10.1109/ICIP.2017.8296389