

# Structure and Infrastructure Engineering

## Maintenance, Management, Life-Cycle Design and Performance

ISSN: (Print) (Online) Journal homepage: [www.tandfonline.com/journals/nsie20](http://www.tandfonline.com/journals/nsie20)

## Creating a first-pass algorithm for corrosion assessment in bridge inspections using machine learning and UAV-collected imagery data

Hana Herndon & Iris Tien

To cite this article: Hana Herndon & Iris Tien (04 Mar 2025): Creating a first-pass algorithm for corrosion assessment in bridge inspections using machine learning and UAV-collected imagery data, Structure and Infrastructure Engineering, DOI: [10.1080/15732479.2025.2470857](https://doi.org/10.1080/15732479.2025.2470857)

To link to this article: <https://doi.org/10.1080/15732479.2025.2470857>



Published online: 04 Mar 2025.



Submit your article to this journal [↗](#)



View related articles [↗](#)



View Crossmark data [↗](#)



# Creating a first-pass algorithm for corrosion assessment in bridge inspections using machine learning and UAV-collected imagery data

Hana Herndon and Iris Tien

School of Civil and Environmental Engineering, Georgia Institute of Technology, Atlanta, GA, USA

## ABSTRACT

Unmanned aerial vehicles (UAVs) have the potential to reduce bridge inspection time and cost while increasing safety. However, UAV-collected field data has inherent properties that complicate damage assessment. In this article, the authors integrate UAV-collected imagery data with automatic defect detection to create a novel first-pass bridge inspection algorithm, which aims to conduct an initial corrosion assessment to determine if further inspection is needed. The authors use UAV-captured images of bridges near Atlanta, Georgia, USA, to create a dataset representative of bridge inspections, including the presence of chaos and misleading objects. The proposed methodology integrates deep learning methods (fully convolutional network (FCN)) to remove natural elements in the image background that resemble corrosion, image processing techniques to quantify texture and reduce lighting effects, and unsupervised learning (K-means) for corrosion segmentation. Experimental results show that the K-means algorithm outperforms other segmentation methods, including image thresholding and deep learning, with a recall of 0.78 and mIoU of 0.72 on UAV-collected field data. Thus, the newly developed method is a promising tool to improve the efficiency and safety of bridge inspections by reducing the number of full inspections conducted on structurally sound bridges.

## ARTICLE HISTORY

Received 27 May 2024  
Revised 19 October 2024  
Accepted 21 November 2024

## KEYWORDS

Bridge inspection; computer vision; image-based corrosion assessment; machine learning; structural health monitoring; unmanned aerial vehicles

## 1. Introduction

Corrosion is a destructive process, characterized by an electrochemical interaction between a metal and its environment (Javaherdashti, 2000). The implications of this process are particularly concerning for civil infrastructure because it leads to significant deterioration, decreases the mechanical properties of the metal, and poses a risk of significant failures such as collapse (Nash et al., 2022). Corrosion is a critical degradation mechanism to consider for bridges in particular as it is one of the most common causes of bridge failure, along with scour and construction mistakes (W. Lin & Yoda, 2017). It is essential to address corrosion in these structures to maintain their integrity and safety. For engineers to make well-informed decisions regarding repairs and reconstruction, they must have a comprehensive understanding of every asset's condition, including the damage induced by corrosion. The effectiveness of prioritization, repair, and reconstruction decisions in bridge asset management relies on conducting rigorous, accurate, and comprehensive bridge inspections. However, existing bridge inspection methods are time-consuming, expensive, and highly disruptive to traffic. They require hours of manual labor and road closures while an expert inspects the structural site (Abdallah et al., 2022; Jahanshahi et al., 2009). In addition, they often put inspectors at risk (J. J. Lin et al., 2021; Tien & Herndon, 2023).

The advent of affordable unmanned aerial vehicles (UAVs) that can carry cameras close to a structure has provided a potential means to improve the time- and cost-efficiency, as well as safety, of bridge inspections. Recent surveys show that an increasing number of state Departments of Transportation (DOTs) are interested in using UAVs for bridge inspections because they are non-contact, time-efficient, and cost-efficient (Jeong et al., 2020; Tien & Herndon, 2023). These UAVs may be equipped with one or multiple sensors, including a super-pixel camera or an infrared camera. Additionally, with the increased interest from DOTs, some UAVs have features that improve their ability to conduct bridge inspections, such as 180-degree camera rotation and obstacle avoidance sensors that reduce the reliance on GPS (Dorafshan et al., 2018; Tomiczek et al., 2019).

To facilitate the bridge inspection process further, improvements in computer vision and machine learning algorithms combined with increasing computer power have opened the door for research into automated corrosion detection and assessment from visual imagery (Nash et al., 2022), which also has the potential to improve the objectivity and accuracy of bridge inspections. Corrosion assessment tasks that can be conducted using computer vision and machine learning algorithms consist of three types: detection, localization, and segmentation. Detection algorithms determine if corrosion exists in the image without specifying its location or extent. Localization algorithms, on the other

hand, pinpoint the location of the defect. Finally, segmentation algorithms determine the size and location of the defect, providing the most information for inspectors.

In this article, the authors create a novel first-pass bridge inspection algorithm that integrates deep learning methods, image processing techniques, and unsupervised learning approaches to segment corrosion in bridge inspection images. This approach aims to segment surface corrosion on exposed steel members, and can be applied to any bridge with exposed steel girders or piles. The authors use UAVs to collect images of bridges near Atlanta, Georgia, USA, creating a dataset that is representative of field bridge inspection conditions. This dataset is characterized by variability in lighting and features in the images, and contains numerous natural elements outside of and surrounding the bridges that mimic corrosion in texture and color, which present challenges for many algorithms.

The article introduces an innovative approach to be used as a first-pass bridge inspection to segment corrosion and compares corrosion assessment algorithms on the same dataset. A key intellectual contribution of this paper is the comparison of multiple approaches on a single UAV-collected dataset, demonstrating the performance of state-of-the-art methods and allowing DOTs to evaluate which algorithms are most suitable for UAV-aided inspections. Additionally, the proposed methods are applied to the widely-used COCO Bridge dataset (Bianchi & Hebdon, 2021a) and another UAV-collected dataset with less chaotic elements to evaluate their generalizability. The practical application of these algorithms to datasets that are representative of real-world UAV-collected inspection data illustrates how UAVs, computer vision, and machine learning could best be integrated into the bridge inspection process.

This paper is organized as follows. Section 2 discusses previous research and background on corrosion assessment using visual imagery. Section 3 describes the proposed methodology for a first-pass bridge inspection. Section 4 presents the results of the methodology applied to field-collected bridge inspection data. Multiple accuracy metrics are provided and the results are evaluated compared to outcomes using other methods. The main conclusions of this paper are summarized in Section 5.

## 2. Background

Most early studies use image processing to detect corrosion in images without localizing or segmenting the defects (Itzhak et al., 1981). While this approach requires the least amount of computational time and complexity, it does not provide comprehensive data about the damage that is important for bridge inspections. Some researchers have used image processing techniques such as texture thresholding and color thresholding to localize and segment corrosion in images (Bonnin-Pascual & Ortiz, 2014; Khayatazad et al., 2020). However, they found that these methods can be prone to false positives and do not perform well in situations with misleading objects or nonuniform lighting, which are characteristic of bridge inspections. Additionally, the

parameters in image processing approaches must be fine-tuned by trial-and-error, which is time consuming (Khayatazad et al., 2022) and not generalizable across applications.

To mitigate these limitations, researchers have investigated the use of machine learning algorithms for corrosion assessment. Son et al. (2014) evaluated six different machine learning approaches for corrosion assessment: support vector machine (SVM), back-propagation neural network (BPNN), J48 decision tree, Naive Bayes, logistic regression, and k-nearest neighbor (KNN). They found that SVM, J48, and KNN all perform with precision and recall above 0.96, with J48 requiring the shortest test time. However, these methods have also been found to be prone to false positives in environments with misleading objects or nonuniform lighting (Bonnin-Pascual & Ortiz, 2014; Khayatazad et al., 2020).

More recently, researchers have investigated the use of deep learning for corrosion segmentation. Deep learning can be advantageous because these algorithms automatically learn features that must be handcrafted in traditional machine learning algorithms and have been shown to perform better on complex datasets. Atha and Jahanshahi (2018) tested multiple convolutional neural network (CNN) architectures for corrosion detection, finding the best one to work with an F1 score of 0.96. Fondevik et al. (2020) segmented corrosion on images of bridges collected by human inspectors with a mean intersection-over-union (mIoU) of 0.78. Bianchi and Hebdon (2021b, 2022) segmented multiple classes of corrosion on images collected by human inspectors with a weighted F1 score of 0.88. Nash et al. (2022) segmented corrosion on images of pipes in industrial yards with an F1 score of 0.84. Zhou et al. (2022) used a CNN to localize corrosion in images, then used image processing to segment the corrosion in the detected bounding boxes. This methodology worked with an F1 score of 0.96. Finally, Rahman et al. (2021) used a semi-autonomous labeling system to create a dataset of 10,000 images that they used to train the CNN DeepLab V3, which performed with an F1 score of 0.81.

While these findings are promising, the majority of them use data that fails to capture the conditions of bridge inspections, which are typically characterized by extensive chaos and noise, e.g. with misleading objects in the image, such as vegetation or graffiti, and nonuniform lighting. Furthermore, they lack validation on datasets of actual bridges, i.e. field-collected datasets, and datasets collected by UAVs, which raises concerns about their generalizability across datasets and applicability in UAV-aided bridge inspections. The investigation by Forkan et al. (2022) highlights this issue. While the algorithm by Atha and Jahanshahi (2018) performed with an F1 score of 0.96 on their dataset, which included photos of corrosion on random objects taken by handheld cameras, its performance dropped to an F1 score of 0.56 when applied to a dataset of images of telecommunications towers taken by UAVs. This study shows the need to subject deep learning methodologies to rigorous testing across diverse datasets that

include the task at hand. For UAVs and automatic damage assessment methodologies to be valuable in bridge inspections, they must be developed and tested with images of bridges within the actual inspection environment. Without such validation, these algorithms are likely to yield unreliable or misleading results.

Additionally, image-based corrosion assessment methods fall short in providing subsurface information, which is crucial to consider when making informed decisions about bridge repair and reconstruction. Therefore, even if an image-based methodology performs flawlessly, inspectors must conduct further analysis to assess the severity and structural impact of corrosion. Considering this and based on the results presented in this study, the authors propose that the best approach for UAV-aided bridge inspections is as a ‘first-pass’ inspection. This involves using UAVs, paired with accurate and comprehensive defect detection and characterization, to conduct an initial assessment of a structure by segmenting all visible surface corrosion. This will determine if the bridge needs further inspection. The methodology is particularly aimed at detecting uniform surface corrosion and is not suitable for identifying corrosion located beneath the surface of the structure, such as corrosion of rebar before concrete spalling. Figure 1 shows examples of corrosion that can and cannot be detected using this approach. If damage is identified, a full inspection should be completed. However, in cases where no damage is detected, a full inspection becomes unnecessary.

The proposed approach can be applied to any bridge with exposed steel, even if it has been painted or metalized, but has been tested on simple span bridges with exposed steel girders and piles covered with green anti-corrosion paint. This approach is not applicable to reinforced concrete bridges, as it cannot detect corrosion of fully covered rebar. In the state of Georgia alone, there are 3,495 bridges with exposed steel for which this approach can be used (*LTBP InfoBridge, n.d.*). Each of these bridges requires inspection every two years, equating to over four inspections per day. In the USA, there are 171,163 bridges with exposed steel girders or piles to which this approach could be applied. By eliminating full inspections on bridges without damage, this methodology will save time and money for agencies and allow inspectors to dedicate more time and resources to evaluate the bridges that are critical.

### 3. Proposed methodology for a first-pass inspection

The proposed methodology consists of two major stages: data collection and image assessment. Data collection consists of image capture using UAVs at the bridge site and image labeling. Note that image labeling is included here as part of the evaluation process of the proposed algorithm. Images collected for new datasets, e.g. in applying the proposed methodology to inspection of a different bridge, do not need to be labeled. The image assessment stage consists of three steps: deep learning to remove the background of the images; image processing, including texture quantification and color space transformation, to create the input



(a)



(b)

Figure 1. (a) Corrosion that can be detected using this approach (from COCO bridge (Bianchi & Hebdon, 2021a)) and (b) spalling in concrete due to corrosion that cannot be detected using this approach (from (Choi et al., 2020)).

features for the algorithm; and unsupervised machine learning to segment corrosion in the images.

Similar to the data labeling, these steps are shown as the description of the proposed algorithm. The algorithm can be automatically applied to new datasets without the need for further model training. It is proposed that this automated approach to data analysis of UAV-collected imagery data will assist in creating efficient decision support for bridge inspections. The full proposed algorithm is shown in Figure 2. Each step is described in more detail in the sections following.

#### 3.1. Data collection and labeling

There are many UAVs available in the market that could be used for bridge inspections. For this application, the Skydio 2+ Enterprise was used. This UAV was selected for its design features that assist with the bridge inspection process, including a vision-based obstacle avoidance system, 180-degree gimbal movement, automatic photo capture, and 3X

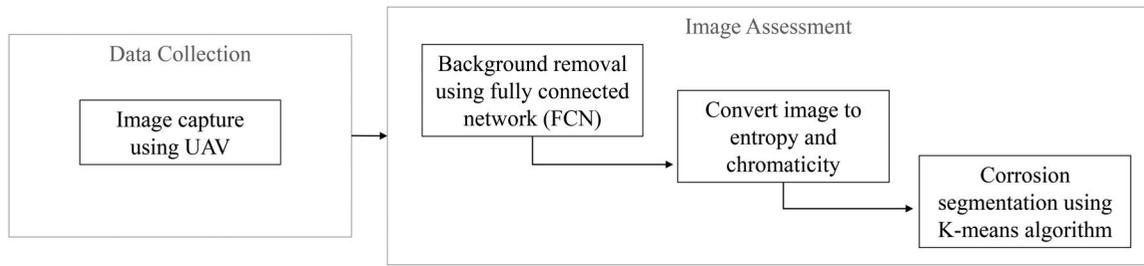


Figure 2. Proposed first-pass bridge inspection algorithm.

zoom. Unlike most other models that use a Global-Positioning-System (GPS)-based obstacle avoidance system, the Skydio 2+ uses small cameras and therefore does not require a strong GPS signal to fly. This gives it the ability to collect data in more locations important for bridge inspections, e.g. underneath a bridge, where a GPS signal may be weak. Additionally, the Skydio 2+ model can avoid obstacles with greater success than GPS-based models because it combines photometric information with scene understanding in real-time (Kesteloo, 2019). However, despite this obstacle avoidance system, as with other models, in practice and as found in the field, it is not capable of avoiding small obstacles that are difficult to see or distinguish such as thin wires or branches that often surround bridge sites.

The 180-degree gimbal movement allows the UAV to take perpendicular photos of the underside and overside of the bridge deck. The automatic image capture and 3X zoom makes it easier to capture high-quality images. All images collected using this UAV have a resolution of  $4056 \times 3040$  pixels, and the automatic image capture feature allows inspectors to collect high-resolution images without pausing the flight, increasing the efficiency of the inspection (Tien & Herndon, 2023).

Although the UAV has a 3X zoom feature, not all the images in this dataset utilized it, resulting in some images with small regions of corrosion and low visibility. These images were retained in the dataset because they represent the kinds of images that may be captured using other UAVs, making it valuable to assess algorithm performance on them. However, the study also tested algorithm performance after removing these less ideal images, to evaluate how the algorithms perform when more optimal images, i.e. more zoomed images with larger regions of corrosion, are used.

Data labeling was completed using the Computer Vision Annotation Tool (CVAT). CVAT is a free, open-source, web-based annotation tool that allows users to label data for all tasks of supervised machine learning (Sekachev et al., 2020). For this inspection, polygon labels were used, which create a pixel-level label in the image and allow for corrosion segmentation. These labels are recommended over bounding box labels because they provide information on the location and extent of corrosion in the image. In applying the proposed methodology to a new dataset, such as imagery data for another bridge, the trained algorithm can be applied directly to the new dataset without creating new labels. In this study, the labels are used to evaluate the

proposed methodology, but are not used to train the corrosion segmentation algorithm.

### 3.2. Image processing to create input features of the algorithm

#### 3.2.1. Texture

This step focuses on processing of the images to create input features of the algorithm based on characteristics specific to corrosion. The surface roughness of steel is a telling visual feature of corrosion. Therefore, the texture in the image can be quantified to assess corrosion. In this work, values of entropy are used to quantify texture in the image with implementation using the skimage package in Python. First, because textural features are computed on single-layer images, the image is converted to greyscale. The grey level indicates the brightness of a pixel; for example, a grey level of zero means that the pixel absorbs all light and appears black (Khayatazad et al., 2022). Once the image is converted to greyscale, the histogram is equalized using `skimage.exposure.equalize_hist` to enhance the image's contrast by spreading out the intensity range of the image. This facilitates separating pixels based on greyscale values and therefore texture quantification as well (Abdullah-Al-Wadud et al., 2007). This study finds that the performance of the texture-based assessment improves after histogram equalization.

Then, the entropy of the images is calculated with a radius of 5 using `skimage.filters.rank.entropy`. Entropy is a statistical measure of randomness, which measures the information or 'surprise' associated with a pixel being a certain value. Equation (1) is used to compute the entropy of the pixels in an image, where  $H$  is the entropy of pixel  $X$ ,  $p(x)$  is the probability that pixel  $X$  has intensity  $x$ , and  $n$  is the number of pixels in the specified surrounding region:

$$H(X) = - \sum_{i=1}^n p(x_i) \log p(x_i) \quad (1)$$

A likely event has less entropy than an unlikely event. If a value is highly likely, it is not surprising and does not provide researchers with as much information as an unlikely event. In the case of texture quantification, the event in question is the value of a pixel. In textured regions of the image, the likelihood of a pixel being a specific value is low, so the entropy in that region is high. Meanwhile, low entropy regions correspond to smooth regions of the image. The entropy values are normalized to be between 0 and 1 before being input to the algorithm. Figure 3 presents the

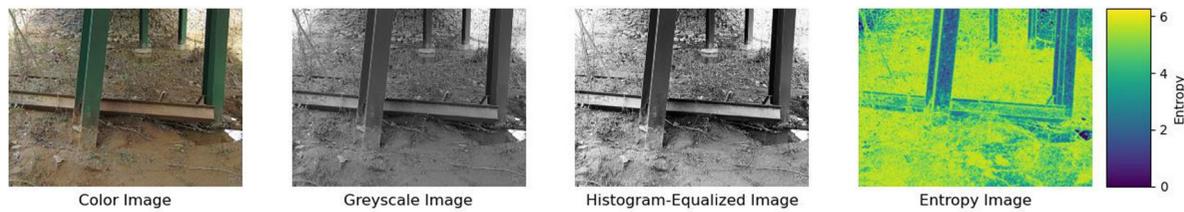


Figure 3. Color (a), greyscale (b), histogram-equalized (c), and entropy (d) of an image from the field-collected dataset.

original, greyscale, histogram-equalized, and entropy of an image from the dataset.

### 3.2.2. Color

Because corrosion on steel has a distinctive red-brown color, color is a potential feature that could be used to segment corrosion from visual imagery. However, an issue with color detection is that it depends on the illumination of the image in addition to the visual characteristics of the object (Khayatazad et al., 2022). Choosing an adequate color space can alleviate this problem and improve the use of color for corrosion assessment (Atha & Jahanshahi, 2018). In the standard red-green-blue (RGB) color space, colors are created by combining different values of red, green, and blue, with each pixel represented by three values ranging from 0 to 255. When all three colors are at their maximum, the result is white, while zero values for all three colors produce black. Therefore, in this color space, white regions have high red values, which does not effectively address the illumination problem for corrosion segmentation.

The hue-saturation-value (HSV) color space has been developed to simulate how humans perceive color rather than combining three primary colors. The hue is the primary channel and takes a value between 0 and 360. Saturation represents the 'grey-ness' of a color and takes a value between 0 and 1. Lastly, value represents the relative lightness of a color and varies between 0 and 1. In previous studies, corrosion was segmented with better accuracy in HSV images than in RGB images (Bondada et al., 2018; Petricca et al., 2016).

Lastly, the  $L^*ab$  color separates the light ( $L^*$ ) values from the red-green (a) and blue-yellow (b) values, thus separating the lightness of a pixel from its chromaticity. This reduces the effects of varying or nonuniform lighting on the results of corrosion segmentation. The lightness values range from 0 to 100, where a value of 0 corresponds to black and a value of 100 corresponds to white. In the a layer, negative values correspond to green and positive values correspond to red. In the b layer, negative values correspond to blue and positive values correspond to yellow. Previous studies have found corrosion to be segmented with better accuracy in  $L^*ab$  images than in RGB and HSV images (Khayatazad et al., 2022). This work has found similar results; therefore, the  $L^*ab$  color space is used in this study. The images were converted using `skimage.color.rgb2lab`. Figure 4 shows an image from the field-collected dataset in this study shown in the RGB (original), HSV, and  $L^*ab$  color spaces.

### 3.3. Background removal

Initial processing on the images found that the background in this dataset is too chaotic to directly conduct image segmentation for corrosion. There is significant natural material in the background that resembles corrosion in texture and color, such as many textured thin brown branches. As a result, there are many false positives because corrosion is predicted to be in the background of the images rather than on the bridge. Therefore, it was determined that the background needs to be removed from the images in this dataset before further processing and analysis.

In previous works, the background did not resemble corrosion as closely and did not produce as many false positives. For example, the images were taken of pipes in industrial plants (Nash et al., 2022), where the background was blue sky or concrete, or taken by human inspectors and therefore closer to the bridge and the background was not shown (Khayatazad et al., 2022; Rahman et al., 2021). However, for UAV-assisted bridge inspections, with images taken by UAVs and in rural or vegetated areas with plants and foliage, the background is difficult to avoid completely and often closely resembles corrosion. Therefore, removing the background is a necessary step in the methodology to reduce false positives.

In this study, a pre-developed deep learning CNN architecture fully convolutional network (FCN) trained on images in the dataset is used to remove the background. Although there are open-source background removal algorithms available, they are not trained on images of bridges and therefore do not perform well on the bridge inspection dataset. The PyTorch implementation of FCN was used to train the algorithm on the collected and labeled data. Once the algorithm is trained, it can remove the background from the images quickly and accurately. A perfect performance would be a mean-intersection-over-union (mIoU) of 1.00 and a loss of 0.00. The final mIoU of this model was 0.925 on the validation set and the final loss was 0.120. These results suggest that this model can remove the background from images of bridges well, albeit not perfectly.

In this methodology, the deep learning background removal method and a perfect case scenario using the ground truth labels are used. In practice, using the ground truth labels is not a practically implementable method to remove the background because it is equivalent to manually removing the background, decreasing the efficiency of the process. Here, this method is tested to explore how the algorithms may work on photogrammetry models or images as deep learning background removal algorithms become more robust and accurate through more diverse training.

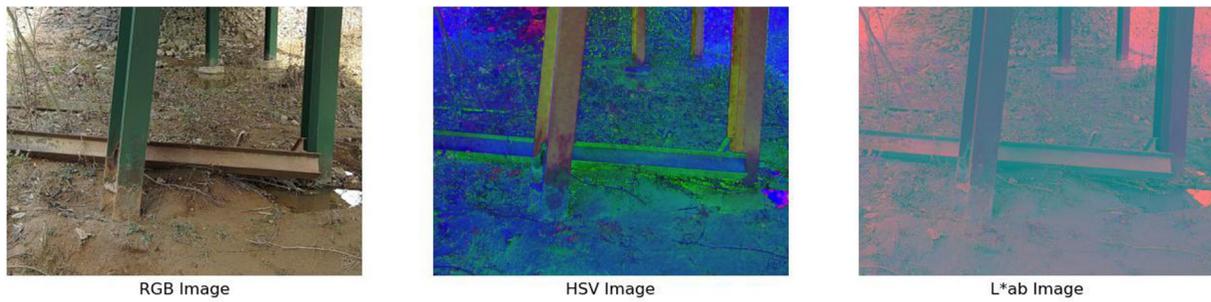


Figure 4. RGB (a), HSV (b), and L\*ab (c) representations of an image from the dataset.



Figure 5. Image from the dataset with the background intact (a), removed using FCN (b), and removed using the ground truth label (c).

Figure 5 depicts an image from the dataset with the background removed using the FCN and the ground truth label.

### 3.4. K-means algorithm

With the images processed and background removed, an unsupervised machine learning approach is used for corrosion segmentation. The K-means algorithm is a standard unsupervised machine learning algorithm that is used to separate data into  $k$  different clusters through an iterative, converging process where each data point belongs to only one group (Na et al., 2010). This algorithm aims to make the intra-cluster points as similar as possible while keeping the clusters as different as possible. Because of the iterative and self-updating nature of the algorithm, it is potentially more accurate than simple thresholding using color or texture values as in previous works for corrosion detection (Bonnin-Pascual & Ortiz, 2014; Petricca et al., 2016). In this work, the sci-kit learn implementation of K-means in Python was used. The similarity metric used was Euclidian distance  $E$ , defined in Equation (2), where  $x$  represents each data point, or in this case, each pixel in the image;  $k$  indicates the number of clusters; and  $C_i$  is the  $i$ th cluster:

$$E = \sum_{i=1}^k \sum_{x \in C_i} (x - x_i)^2 \quad (2)$$

An additional challenge in assessing corrosion on bridges from UAV-collected images is that the data is often unbalanced. That is, corrosion is often represented in a much smaller portion of the image compared to the non-corroded parts of the image. Unsupervised learning algorithms, such as K-means, can be better suited for unbalanced data because they do not use loss functions to find patterns in

the data, but instead look at similarities of the input parameters. For unbalanced data, the loss function can be minimized sufficiently without achieving the goal of the analysis. For example, in this field-collected dataset of bridge inspections, on average around 2% of the pixels in a given image represent corrosion. A supervised learning algorithm can achieve a loss of 2% if the entire dataset is labeled as non-corrosion. This outcome does not serve to support assessment of corrosion on bridges. In addition, using an unsupervised learning algorithm means that inspectors can apply this methodology on new bridges without needing to label the data, which is a tedious and time-consuming task.

The images used in this study with the K-means algorithm had one layer of normalized entropy, one layer of normalized red-green values as in the L\*a\*b color space, and one layer of normalized blue-yellow values as in the L\*a\*b color space. Different values of  $k$  ranging from one to six were tested, and it was found that specifying  $k=4$  yields the best results for corrosion segmentation. The cluster with its center having the highest average values—that is, where a-chromaticity was the most red, b-chromaticity was the most yellow, and texture entropy was the highest—was identified as the cluster representing the corrosion class.

## 4. Application to field-collected bridge inspection data and results

### 4.1. Details of bridge geometry and environment

To train and evaluate the proposed algorithm, extensive field data was collected at bridges in two locations owned and operated by the Georgia Department of Transportation (GDOT) near Atlanta, Georgia, USA. One bridge is located

in Douglasville, Georgia, approximately 30 miles west of Atlanta, and the other is located in Calhoun, Georgia, approximately 70 miles northwest of Atlanta. The bridges have steel piles and girders with concrete abutments and decks. Both bridges were reported by GDOT to have corrosion on the piles, and the bridge in Douglasville was reported to have corrosion on the girders (Tien & Herndon, 2023).

Both bridges are located in rural areas, crossing creeks and surrounded by numerous trees and bushes, limiting where the UAV can fly and what sections of the bridge it can capture, and creating chaotic backgrounds for the images with many misleading objects. Figure 6 shows images captured at the bridge sites, highlighting the vegetation, thin branches, and types of corrosion present in this dataset. This environment is characteristic of bridge inspection conditions; for automatic or semi-automatic bridge inspections to be useful in the field, they must perform well in this type of environment. A total of 1,893 UAV-collected images were captured at these locations. A total of 767 images were labeled using CVAT, with polygon labels to create pixel-level labels and allow for image segmentation. Labeled images were representative of the overall set of collected images.

## 4.2. Results

This section investigates the performance of the proposed approach for corrosion assessment in UAV-collected images of bridge inspections. Numerical performance on all datasets is shown in Table 1. After training the background removal algorithm, all steps of the algorithm were conducted on a Dell XPS computer with a 12th Generation Intel® Core. Image loading, background removal, and conversion to the input parameters took approximately 10 min and training the K-means algorithm required approximately 11 min on this personal machine. The visual performance of K-means for corrosion segmentation on images with the background removed using FCN are shown in Figure 7.

As shown in Table 1, the methodology performs with good accuracy and mIoU, but it has low precision and recall. The mIoU is comparable with previous research, i.e. 0.72 compared to 0.78 in Fondevik et al. (2020), but the

Table 1. Performance of the K-means algorithm on the various datasets.

Dataset	mIoU	Accuracy	F1 score	Precision	Recall
UAV-Collected (FCN bg removal)	0.7834	0.7976	0.0951	0.0592	0.5129
UAV-Collected	0.7167	0.7244	0.1037	0.0612	0.7772
COCO Bridge	0.6008	0.6300	0.4491	0.3972	0.7274
Combined Dataset	0.6800	0.7336	0.2692	0.2167	0.7114



(a)



(b)

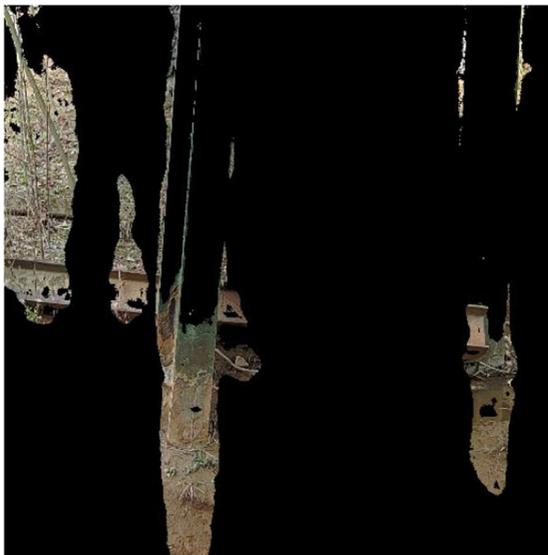


(c)

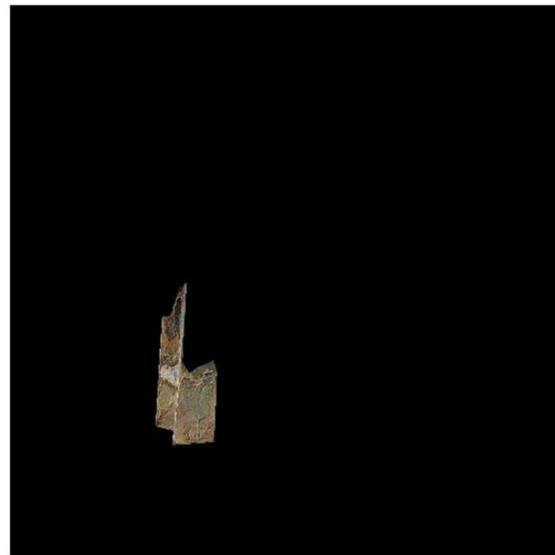


(d)

Figure 6. Images captured at the bridge sites: (a) image captured in Douglasville, Georgia, with vegetation and thin branches present throughout. (b) Image of surface corrosion on a girder on the bridge in Douglasville. (c) Image captured in Calhoun, Georgia, with vegetation present throughout. (d) Image of surface corrosion on a pile of the bridge in Calhoun.



Results of K-means with Background Removed using FCN

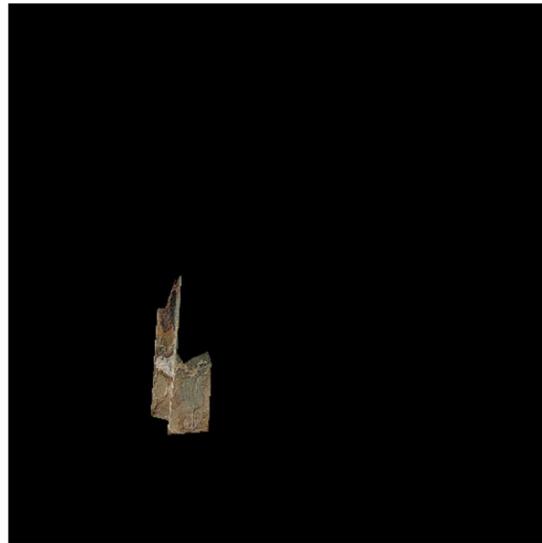


Ground Truth Corrosion Inference

Figure 7. Results of K-means with background removed using FCN (a) and the ground truth corrosion inference (b).



Results of K-means with Background Removed using Ground Truth



Ground Truth Corrosion Inference

Figure 8. Results of K-means with background removed using the ground truth (a) and the ground truth corrosion inference (b).

recall is too low to be useful in bridge inspections. A recall of 0.51 suggests that 49% of corroded regions on the bridge would go unnoticed if this methodology were used in a bridge inspection. Ideally, the recall would be 1.0, particularly for bridge inspection applications where false negatives should be limited for the safety implications of missed areas of corrosion in inspections. Therefore, as is, this is not a viable method to aid in bridge inspections. To determine the performance of this methodology in a best-case scenario, the background of the images is removed using the ground truth labels before K-means clustering for segmentation. The visual performance of this method are shown in Figure 8.

The performance improves to varying degrees when the background is removed using the ground truth label. Importantly, the recall increases significantly, from 0.51 to 0.77, or 50.1% percent. Surprisingly, however, the precision, accuracy, and mIoU decrease. This result is more surprising

when inspecting the inferences visually because it appears that more of the image was correctly inferred. This result could be influenced by several factors. Firstly, in situations with a class imbalance between labels, metrics might not accurately reflect the algorithm's ability to correctly identify the minority class, which is corrosion in this case. Additionally, the ground truth label used for corrosion may have limitations or inaccuracies, such as inexact labeling of corrosion on the images at thresholds between corroded and uncorroded portions, which introduces errors into the algorithm and negatively impacts the performance metrics. Lastly, the complexity of the task itself can pose challenges for pixel-level segmentation metrics, which may not provide a holistic view of the algorithm's performance because they focus on the classification of individual pixels rather than the overall semantic understanding of the image.

The accuracy and mIoU values are in the 0.7 range, with a slight decrease from the deep learning scenario to the

perfect case scenario. Ideally, these values would be higher, but they are close to the accepted human accuracy for segmentation in images, which is 0.8 (Nash et al., 2022) and the benchmark used in other segmentation research. Therefore, although the accuracy has decreased, the notable increase in recall and segmentation improvements by visual inspection suggest that as background removal improves, corrosion segmentation using K-means also improves.

To assess this algorithm's generalizability across datasets, its performance was also assessed on the COCO Bridge dataset developed by Bianchi and Hebdon (2021a). The images, collected by human inspectors at the Virginia Department of Transportation (VDOT), are more ideal compared to the UAV-collected images. For comparison, the corrosion class comprises 21.0% of this dataset, compared to 1.7% in the original UAV-collected dataset. All levels of corrosion (good, fair, and poor) labeled in the COCO Bridge dataset are considered as corrosion for this study. The background was not removed from these images because the human-collected images in this dataset are less chaotic than UAV-collected ones. The performance metrics after converting the image to an entropy layer, a-layer, and b-layer, then using K-means to segment corrosion are shown in Table 1.

As shown, the proposed algorithm has a similar recall on this COCO Bridge dataset compared to the original UAV-collected dataset. The mIoU decreases slightly, suggesting that it is more difficult for the algorithm to segment the background class when it is less prominent. The most significant performance change is the precision, which increases from 0.0612 to 0.3972, subsequently raising the F1 score to 0.4491. These results suggest that this methodology's ability to segment all regions of corrosion, as represented by the recall, remains relatively consistent across datasets. Additionally, in images with larger and clearer portions of corrosion, the algorithm's ability to remove uncorroded regions improves.

To further evaluate the proposed algorithm's ability to perform on higher-quality UAV-captured images, particularly those with larger portions of corrosion, it was tested on a third dataset. This dataset consists of UAV images captured according to best practices for UAV data collection found in this study combined with the COCO Bridge validation dataset. The UAV best practices data collection guidelines include taking photos in closer proximity to the bridge and making extensive use of the zoom feature to ensure high detail in the images and larger regions of corrosion. Images that did not meet these criteria were excluded from the dataset. The corrosion class comprises 10.7% of this dataset, almost 10 times that of the original UAV-collected dataset. This dataset provides an additional opportunity to assess the algorithm's generalizability and evaluate its performance on UAV-collected images that are simultaneously representative of bridge inspection conditions in the field and more suitable for algorithmic inference.

Again, the recall remains consistent when this algorithm is applied to the third dataset, highlighting the robustness of the proposed methodology in segmenting all regions of

corrosion during a first-pass bridge inspection across various conditions. While the precision is lower for this dataset than the COCO Bridge dataset, it is significantly higher than that for the original UAV-collected dataset. This improvement in precision further supports the idea that, in images with larger portions of corrosion, the algorithm's ability to remove uncorroded regions improves, enhancing its overall performance in real-world scenarios.

For comparison, the state-of-the-art for corrosion segmentation is reported to achieve an F1 score of 0.96 (Zhou et al., 2022). However, this research uses images of crane structures, which are not surrounded by vegetation and therefore are not in environments representative of bridge inspections. The state-of-the-art for corrosion segmentation on civil structures is reported to perform with a weighted F1 score of 0.88 (Bianchi & Hebdon, 2022). However, this methodology is not tested on images collected by UAVs, and are therefore likely to suffer decreases in performance when used on an external dataset as in other deep learning methodologies (Atha & Jahanshahi, 2018; Forkan et al., 2022). Results of applying previously proposed high performing methods (i.e. a deep learning method with a reported F1 score of 0.88 in Bianchi and Hebdon (2022)) to the field-collected dataset of this study are presented in the following section. Results show the degradation in performance when applied to external UAV-collected data from the field.

### 4.3. Comparison with other methods

The proposed algorithm for a first-pass bridge inspection is created based on the highest performing results compared to other computer vision methods including image processing techniques, such as texture thresholding and color thresholding, and deep learning approaches. These methods do not perform as well as K-means for various reasons. The following sections describe their varying limitations and demonstrate the challenges of adapting existing techniques from the literature to new external datasets.

#### 4.3.1. Texture thresholding

To investigate texture quantification for corrosion assessment, two methods are investigated: entropy, as discussed in Section 3.2, and edge detection, also known as gradient operators. Edge detection methods use convolution filters to show where pixel values change rapidly. Textured portions have more pixel variation, and therefore have more edges than smooth ones. It was found that these two yield similar results, so entropy was selected in the proposed algorithm for its decreased computational burden.

Once the texture is quantified using entropy in the images, a threshold is used to segment the image. This threshold was determined by visual analysis. First, a group of four test images, which allowed complete and comprehensive visual analysis of the images, were segmented using different entropy thresholds. In each image, ten thresholds at 10% intervals of the maximum entropy were used to see

at which threshold the corrosion was fully separated into one cluster. Visual analysis showed that a threshold of 65% of the maximum entropy resulted in the best segmentation of corrosion. Because of the foliage present at the data collection sites, a lot of texture is detected in the background of the image. To reduce the number of false positives and focus the corrosion inferences on the bridge, the background is removed using an FCN and ground truth labels, as described in Section 2.3, and then entropy is used to segment corrosion in the images.

This methodology (i.e. automated background removal and texture thresholding) works with high accuracy (0.8527) and mIoU (0.8390), but other metrics such as precision (0.0818) and recall (0.5050) are low. This is because the amount of corrosion in the images is small, so when the negative space is inferred accurately, the overall accuracy is very high even if the corrosion is not inferred well. On the COCO Bridge and combined datasets, these metrics improve, with recall surpassing 0.6 in both cases. However, this still does not reach the performance of K-means, nor is it sufficient for use in UAV-aided bridge inspections.

#### 4.3.2. Color thresholding

Color thresholding was tested in three color spaces: Red-green-blue (RGB), hue-saturation-value (HSV), and lightness-a-b ( $L^*ab$ ). To segment the images, the 25th and 75th percentiles of each layer were found, and then the distance between these percentiles and each pixel were calculated. If the values of a pixel were closer to the 25th percentile, that pixel was classified into the first cluster, and if the values were closer to the 75th percentile, the pixel was classified into the second cluster. This thresholding method is used to account for the fact that each pixel has multiple values, rather than in the texture scenario, where each pixel has only one.

In the RGB color space, color thresholding did not appear to segment corrosion in the images, but rather separated the light space from the shadows. In the HSV color space, to follow previous studies methodologies, the value layer is removed and the image is segmented using only the hue and saturation, which is supposed to reduce the impact of lighting (Bondada et al., 2018; Petricca et al., 2016). However, for the dataset in this study, this approach resulted in the image being segmented not for corrosion but based on grey areas and colorful areas. In the  $L^*ab$  color space, the lightness layer is removed and the image is segmented based on the values in the color layers. This color space performs the best; the clusters appear to be separated by color rather than lightness or saturation and the corroded patches are more fully segmented into one cluster rather than being split between both of them. Therefore, the proposed methodology analyses images in the  $L^*ab$  color space.

The color thresholding was also tested with the background removed in an attempt to reduce false positives. However, as with texture thresholding, color thresholding on its own results in performance with high accuracy but low precision and recall. Again, this is possible because the

majority of the image represents uncorroded areas. In the COCO Bridge and combined datasets, there are large improvements to precision and recall, but again, not to a level that is sufficient for the desired corrosion segmentation of this study.

#### 4.3.3. Gaussian Mixture modeling

Gaussian Mixture Models (GMMs) are another standard unsupervised machine learning approach. GMMs assume that all data points are generated from a mixture of a finite number of Gaussian distributions with unknown parameters. The Gaussian parameters are estimated from the training data using the iterative Expectation-Maximization (EM) algorithm (Reynolds, 2009). Although GMMs typically employ soft clustering—determining the probability that a point belongs to a cluster—this application uses hard clustering, where each point is assigned to the cluster with the highest probability. This approach is consistent with the other methods tested in this study, where each pixel belongs with certainty to one class. With the high performance of K-means, GMMs were also tested on the datasets to evaluate whether another clustering approach could yield improved results.

To facilitate comparison, the data was pre-processed in the same way as for the K-means algorithm, with one layer of normalized entropy, one layer of normalized red-green values as in the  $L^*ab$  color space, and one layer of normalized blue-yellow values as in the  $L^*ab$  color space. With the images processed and the background removed, the sci-kit learn implementation of GMM in Python was used. Different numbers of mixture components were tested, with performance improving until three components were used, beyond which the performance remained unchanged. Using three mixture components, GMM performed poorly on the UAV-collected dataset with a precision of 0.0241 and a recall of 0.2388. The performance improved on the other datasets, but not enough to be useful for bridge inspections.

#### 4.3.4. Deep learning

Deep learning, requiring more computing power but less human intervention, has seen significant advancements in recent years. Convolutional neural networks (CNNs) and Vision Transformers (ViTs) have been used for image classification and recognition (Sookpong et al., 2023; Yao et al., 2019). The biggest advantage of deep learning is that feature extraction is included in the algorithm; as long as there are enough data samples, they learn the appropriate classification features that otherwise must be hand-engineered, as with the texture and color algorithms described above. The elimination of the need of prior knowledge and human effort, in addition to increased computing power, has led to a boom in research into CNNs (Atha & Jahanshahi, 2018).

However, deep learning algorithms, although they might have the same purpose (i.e. corrosion segmentation), are not yet generalizable for use across datasets. For example, a deep learning model recently developed to segment corrosion in UAV-collected images was reported to segment

corrosion with an F1 score of 0.84 (Nash et al. (2022)). However, when this model was used to segment corrosion on the bridge inspection images in this study, the F1 score dropped to 0.24. If these algorithms are to be used on other datasets that they were not trained on, they need to be retrained (either completely or via transfer learning) on sample data from the environment in question. This requires additional labels, which are tedious to create. This section describes the results of CNNs and ViTs applied to the datasets of interest in this study.

**4.3.4.1. CNN trained on site-specific data.** First, an FCN was trained on the field-collected images of bridges from this study to test its ability to predict corrosion. The training, validation, and inference were implemented in PyTorch using 8.5 GB of a graphics processing unit (GPU). After image augmentation, the size of the dataset was 5369, and the model was trained for 50 epochs with a batch size of 5 and a learning rate of  $1e-5$ . With these hyperparameters, training took seven hours to complete. Once training was complete, the model could infer on the dataset almost instantaneously. The final validation loss was 0.0963 and the final validation mIoU was 0.9253.

Although the FCN infers on the validation dataset with low loss and high IoU, closer inspection revealed that it did not segment corrosion in the images at all. In other words, it did not indicate where corrosion is present in the images. Because corrosion is a small portion of the image, the algorithm simply assumes there is no corrosion present on the bridge, resulting in high performance metrics, but not performing the task at hand.

**4.3.4.2. CNN trained on site-specific data and COCO Bridge dataset.** To combat the imbalances in the training data, the following deep learning techniques were trained on both the COCO Bridge dataset developed by Bianchi and Hebdon (2021a) and the UAV-collected data used in this study. Because the images in the COCO Bridge dataset contain larger portions of corrosion, combining the datasets was expected to make it more challenging for the algorithm to achieve low loss without segmenting corrosion. All multi-class labels of corrosion were considered as corrosion for this study. This resulted in a training dataset of 731 images, which were then augmented as in Bianchi and Hebdon (2022), increasing the training set size 3,655 images.

The DeepLabV3 (DLV3) algorithm was trained with the same hyperparameters as those used by Bianchi and Hebdon (2022), where it successfully segmented corrosion in human-collected images of bridges with a weighted F1 score of 0.8860. These hyperparameters included a batch size of 2, 50 epochs, and a learning rate of  $1e-5$ . The training, validation, and inference were also implemented using PyTorch. After training, the final validation loss was 0.1207 and the final validation mIoU was 0.5913. On the entire UAV-collected dataset, this approach performed with a lower recall than previously described methods (0.2559) but higher precision (0.3127), although both values are still low.

This model was then validated on the COCO Bridge dataset to compare it to the original performance. This evaluation yielded a binary F1 score of 0.5582, compared to Bianchi's multiclass weighted F1 score of 0.8857. For a more direct comparison, the weighted F1 score was calculated to be 0.8869. Finally, the dataset was tested on the combined COCO Bridge and best practices UAV-collected field dataset. On this dataset, the recall remains about the same at 0.5421, but the precision increases to 0.6688. However, given the variability in image quality during UAV-aided inspections and the model's inability to consistently achieve high recall, K-means offers a more reliable solution for segmenting corrosion across datasets and inspection conditions.

**4.3.4.3. ViT trained on site-specific data and COCO Bridge dataset.** ViTs are an extension of the transformer architecture, originally developed for natural language processing tasks. ViTs adapt this architecture to computer vision by representing images as a sequence of patches and applying self-attention to learn relationships between them (Sookpong et al., 2023). Recently, these algorithms have emerged as potentially superior algorithms for segmentation, often outperforming CNNs in such tasks (Sookpong et al., 2023). These methods were tested on the combined COCO Bridge and UAV-collected dataset. One ViT model was trained and tested without background removal to evaluate the algorithm's ability to handle data without extensive pre-processing, while another was trained with background removal to explore potential improvements. The datasets were augmented to increase the training set size to 5,848 images and both ViTs were trained with the same hyperparameters, including a batch size of 5, a learning rate of  $1e-5$ , and 60 epochs. The model architecture used was ViT Base 16 (Wu et al., 2020).

The first ViT, trained on images without the background removed, performed with the lowest accuracy (0.6170) and F1 score (0.0115) of all methods tested. The second ViT, which was trained and tested on images with the background removed, achieved a higher F1 score (0.0331). However, this algorithm's performance varied significantly across datasets. On the COCO Bridge dataset, it achieved an F1 score of 0.5045 and a recall of 0.7211. On the dataset combined with UAV-collected images, however, the F1 score dropped to 0.2462, with a recall of 0.3291. The first ViT performed consistently poorly across datasets, with F1 scores below 0.1600 in all cases.

Table 2. Performance metrics of all investigated methods on the full UAV-collected dataset.

Method	mIoU	Accuracy	F1 score	Precision	Recall
K-means	0.7167	0.7244	0.1037	0.0612	0.7772
Texture Thresholding	0.8390	0.8527	0.1236	0.0818	0.5050
Color Thresholding	0.8731	0.8874	0.0817	0.0564	0.2538
GMM	0.6125	0.7783	0.0438	0.0241	0.2388
FCN	0.9253	0.9037	0.0000	0.0000	0.0000
DLV3	0.5879	0.9883	0.2479	0.3127	0.2559
ViT (1)	0.6156	0.6170	0.0115	0.0076	0.3211
ViT (2)	0.9836	0.9837	0.0331	0.0692	0.0271

#### 4.3.5. Comparison of results

Table 2 summarizes the performance of all the methods tested to segment corrosion on the entire UAV-collected validation dataset. The top row provides the results of the proposed methodology using K-means in comparison to results using texture thresholding, color thresholding, FCN, DLV3, and ViT.

In a first-pass bridge inspection, the goal is to detect all corroded regions on the bridge without providing too many false positives. These goals are represented best in the recall and mIoU metrics, which measure the amount of corrosion that is correctly labeled as corrosion and the amount of overlap between the predicted segmentation and the ground truth segmentation, respectively. K-Means segmentation combined with robust background removal is able to segment corrosion in these images with the highest recall, including achieving a recall two to three times higher than that of the best performing deep learning methods. Texture thresholding performs with the second highest recall, though the recall value of 0.51 indicates a large amount of corrosion is not identified. While K-means does not perform with an mIoU as high as values reported in previous research (Atha & Jahanshahi, 2018; Bonnin-Pascual & Ortiz, 2014; Petricca et al., 2016), the relatively high recall value of 0.78 is comparable to the best performing prior studies and importantly, is achieved on actual UAV-collected images of bridges from the field.

To investigate the algorithms' generalizability and performance when applied to different datasets, these methodologies were tested on the COCO Bridge dataset developed by Bianchi and Hebdon (2021a) as well. This dataset consists of human-collected photos and are more ideal (in terms of lighting and background, and with larger portions of visible corrosion) compared to UAV-collected images. To evaluate algorithm generalizability and provide a comparison point on this widely used dataset, performance metrics across all investigated methods are shown in Table 3.

As shown in Table 3, the K-means algorithm has a comparable performance on this dataset as on the UAV-collected one, with a recall above 0.7 and mIoU above 0.6. The precision increased from 0.0612 to 0.3972, increasing the F1 score to 0.4491. When tested on this dataset, DLV3 and ViT(2) perform with higher F1 scores than K-means, with ViT(2) also having a comparable recall to K-means. These findings show the importance of capturing high-quality images—with appropriate lighting, proximity to the structure, and large sections of visible corrosion—and that if these images are available, both K-means and ViTs may be

well suited to aid in bridge inspections. However, such images may not always be available when using UAVs. Additionally, comparing the results from Tables 2 and 3 show that the K-means algorithm has impressive generalizability across datasets and sees performance improvements, particularly in precision, if inspectors can make the effort to capture images with clear views of the bridge components under inspection.

To further evaluate the algorithms' ability to perform on higher-quality UAV-captured images, a third dataset was used for validation. This dataset is the combination of UAV images captured according to best practices and the COCO Bridge validation dataset. The corrosion class comprises 10.7% of this combined dataset (compared to 1.7% in the original UAV-collected dataset and 21.0% in the COCO Bridge dataset). This provides an opportunity to assess the algorithms' generalizability and evaluate their performance on UAV-collected images that are more suitable for algorithmic inference. The performance metrics for this dataset across all investigated methods are summarized in Table 4.

Overall, the performance metrics for all algorithms improve compared to the initial UAV-collected dataset, but are generally worse than when evaluated on the COCO Bridge dataset only. The exception is DLV3, which achieves a high F1 score of 0.6019. However, this is accompanied by a recall that is still significantly lower than that of K-means, which continues to perform with the highest recall and is the only algorithm to achieve a recall above 0.7. This suggests that K-means is the most suitable for use in a first-pass inspection, where segmenting all corroded regions is more important than eliminating non-corroded regions. Additionally, the consistently high recall across datasets suggests that K-means is the most generalizable algorithm and the best suited for handling UAV-collected data. Despite the F1 score improvement, DLV3's lower recall indicates that it is not effective for use in a first-pass inspection. ViT(2)'s performance drops on this dataset, with a recall of 0.3291 and a precision of 0.2648, highlighting that even when best practices are followed during UAV photo collection, this algorithm struggles to handle these images effectively.

#### 4.4. Advantages and limitations of the proposed methodology

One of the main advantages of the proposed approach is that it does not require training labels for corrosion. As K-means is an unsupervised learning algorithm, it separates corrosion from the rest of the image based on the image's

Table 3. Performance metrics of all investigated methods on COCO bridge dataset.

Method	mIoU	Accuracy	F1 score	Precision	Recall
K-means	0.6008	0.6935	0.4491	0.3972	0.7274
Texture Thresholding	0.5740	0.6728	0.3823	0.3505	0.6382
Color Thresholding	0.6641	0.7605	0.4350	0.4944	0.5863
GMM	0.6205	0.6300	0.2394	0.2254	0.3385
FCN	0.6518	0.7899	0.0000	0.0000	0.0000
DLV3	0.6047	0.8586	0.5582	0.5753	0.5421
ViT (1)	0.5869	0.6271	0.1592	0.1833	0.2925
ViT (2)	0.7227	0.7932	0.5045	0.4406	0.7211

Table 4. Performance metrics of all investigated methods on the combined UAV and COCO dataset.

Method	mIoU	Accuracy	F1 score	Precision	Recall
K-means	0.6800	0.7336	0.2692	0.2167	0.7114
Texture Thresholding	0.7178	0.7731	0.2742	0.2234	0.6116
Color Thresholding	0.7723	0.8271	0.2926	0.2642	0.4344
GMM	0.6552	0.6630	0.1308	0.1129	0.3146
FCN	0.8172	0.8930	0.0000	0.0000	0.0000
DLV3	0.6901	0.9795	0.6019	0.6688	0.5472
ViT (1)	0.6772	0.6951	0.0762	0.0840	0.2573
ViT (2)	0.8651	0.8948	0.2462	0.2648	0.3291

characteristics. Therefore, bridge inspectors or agencies will not need to label corrosion in new inspection images to use this methodology, which is a tedious and time-consuming task. Additionally, as shown, this algorithm performs the most consistently across datasets, especially in achieving consistent recall, making it a reliable choice for detecting corrosion on external data. K-means is also better suited for unbalanced data—common in UAV-collected inspection images—compared to supervised learning algorithms. Finally, this algorithm does not require a large amount of computational power to predict corrosion. Once the background is removed, which can also be done with low computational power after training, corrosion prediction can be done on most personal or organizational computational systems.

The primary disadvantage of the proposed methodology is that using K-means for segmentation is better suited for images with larger portions of corrosion. This is shown in the increase in precision as the proportion of corrosion in the image grows. The algorithm does not work on images where there is no corrosion present because it will always segment a portion of the image as corroded. A process may need to be implemented to mitigate this to reduce the number of potential false positives. However, because recall remains consistent across datasets, K-means is still the best option for images with small portions of corrosion among the algorithms tested in this study.

#### 4.5. Enhancing the proposed methodology

While the results of the proposed methodology support its use in a first-pass bridge inspection, there are opportunities to enhance the proposed approach. A possible modification is to use a classification algorithm to determine if there is corrosion present in the image before it is analyzed by K-means for segmentation. Then, all of the images that enter the algorithm will have some level of corrosion for clustering, helping to reduce the number of false positives in the output. The algorithm can also be paired with other machine learning techniques, such as SVM or decision trees as used in Son et al. (2014). Pairing different algorithms has been found to improve the performance of both algorithms in a study by Bonnin-Pascual and Ortiz (2014) and may be able to reduce the number of false positives.

Another alteration to improve this methodology could be adjusting the input parameters for the algorithm. In this methodology, the input features were entropy and the a and b layers from the L\*ab color space. Other methods to quantify the texture, such as the Grey-Level Co-occurrence Matrix (GLCM), dissimilarity, or homogeneity as used in the study by Khayatazad et al. (2022), could be tested. Additionally, multiple texture or color features could be used, rather than just one of each.

It is important to note that this methodology segments corrosion and evaluates the results at the pixel level. While this enables precise results and quantification of performance, this may be a more demanding task than is needed for a first-pass bridge inspection. For example, one pixel

labeled as corrosion surrounded by uncorroded pixels would not be critical for a bridge inspector to identify because corrosion is expected to occur over larger regions than a single pixel. A potential modification to the approach could be to use a bounding box algorithm, such as YOLO V3 used by Zhou et al. (2022), to localize corrosion in the images. However, these algorithms would be subject to the same limitations as other deep learning algorithms, including potentially low generalizability and training difficulty on unbalanced data. Additionally, as this method would localize corrosion but not segment it, K-means could then be used to quantify the corrosion in the bounding boxes to determine if an inspector needs to conduct a full bridge inspection.

## 5. Conclusions

This study investigates image-based corrosion assessment methodologies to create a novel first-pass bridge inspection algorithm. The algorithm includes background removal using FCN, image processing to obtain texture and chromaticity features, and unsupervised K-means to perform corrosion segmentation. In this methodology, the entropy and a and b color channels from the L\*ab color space are used as inputs for the K-means algorithm. The methodology performs with a recall of 0.78 and an mIoU of 0.72.

The authors use UAV-captured images of bridges near Atlanta, Georgia, USA to create a dataset that is representative of bridge inspection conditions. This dataset contains chaos in the images and features several environmental elements (e.g. thin branches, wires, hives) that mimic corrosion in texture and color, which presents challenges for many algorithms. The proposed algorithm is tested on this dataset to reveal its applicability to data that is collected by UAVs in the field, the COCO Bridge dataset developed by Bianchi and Hebdon (2021a) to reveal its generalizability across datasets, and on a combined dataset of COCO Bridge and UAV images collected using best practices to assess its performance on higher-quality UAV-collected images. Results illustrate how growing UAV, computer vision, and machine learning technologies could best be integrated into bridge inspections as an initial corrosion assessment to determine if a bridge needs further inspection. Different computer vision and deep learning algorithms are also investigated, but yield worse results, showing that deep learning methods, even high performing ones, cannot simply be taken from previous research and applied to bridge inspection data.

This study presents an innovative approach to be used as an initial bridge inspection and compares alternative corrosion assessment algorithms applied to the same dataset. However, no technique as of yet works perfectly, and there is room for further improvement in the algorithms investigated in this study to perform at a level high enough to be directly used in bridge inspections. Although manual inspections remain essential for a comprehensive assessment of bridge conditions, a reliable corrosion detection and localization algorithm could significantly reduce the number of full inspections conducted on structurally sound bridges.

This will enable agencies to increase the time and resource efficiency, while improving safety, of bridge inspections.

## Disclosure statement

No potential conflict of interest was reported by the authors.

## Funding

This project was partially funded by the INSPIRE University Transportation Center (UTC). Financial support for INSPIRE UTC projects is provided by the U.S. Department of Transportation, Office of the Assistant Secretary for Research and Technology (USDOT/OST-R) under Grant No. 69A3551747126 through INSPIRE University Transportation Center (<http://inspire-utc.mst.edu>) at Missouri University of Science and Technology. The views, opinions, findings, and conclusions reflected in this publication are solely those of the authors and do not represent the official policy or position of the USDOT/OST-R, or any State or other entity. Additional funding support from the Georgia Department of Transportation through Grant No. RP22-01 is also acknowledged. Author Herndon acknowledges support by the National Science Foundation Graduate Research Fellowship under Grant No. DGE-2039655. Any opinion, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation. Support from the Williams Family Professorship for Author Tien is also acknowledged.

## Data availability statement

Data, models, or code that support the findings of this study are available from the corresponding author upon reasonable request.

## References

- Abdallah, A. M., Atadero, R. A., & Ozbek, M. E. (2022). A state-of-the-art review of bridge inspection planning: Current situation and future needs. *Journal of Bridge Engineering*, 27(2), 03121001. [https://doi.org/10.1061/\(ASCE\)BE.1943-5592.0001812](https://doi.org/10.1061/(ASCE)BE.1943-5592.0001812)
- Abdullah-Al-Wadud, M., Kabir, M., Akber Dewan, M., & Chae, O. (2007). A dynamic histogram equalization for image contrast enhancement. *IEEE Transactions on Consumer Electronics*, 53(2), 593–600. <https://doi.org/10.1109/TCE.2007.381734>
- Atha, D. J., & Jahanshahi, M. R. (2018). Evaluation of deep learning approaches based on convolutional neural networks for corrosion detection. *Structural Health Monitoring*, 17(5), 1110–1128. <https://doi.org/10.1177/1475921717737051>
- Bianchi, E., & Hebdon, M. (2021a). *Corrosion condition state semantic segmentation dataset* (p. 334672716 Bytes) [Dataset]. University Libraries, Virginia Tech. <https://doi.org/10.7294/16624663.V2>
- Bianchi, E., & Hebdon, M. (2021b). *Trained model for the semantic segmentation of corrosion condition states* (p. 803835104 Bytes) [Computer software]. University Libraries, Virginia Tech. <https://doi.org/10.7294/16628668>
- Bianchi, E., & Hebdon, M. (2022). Development of extendable open-source structural inspection datasets. *Journal of Computing in Civil Engineering*, 36(6), 04022039. [https://doi.org/10.1061/\(ASCE\)CP.1943-5487.0001045](https://doi.org/10.1061/(ASCE)CP.1943-5487.0001045)
- Bondada, V., Pratihari, D. K., & Kumar, C. S. (2018). Detection and quantitative assessment of corrosion on pipelines through image analysis. *Procedia Computer Science*, 133, 804–811. <https://doi.org/10.1016/j.procs.2018.07.115>
- Bonnin-Pascual, F., & Ortiz, A. (2014). Corrosion detection for automated visual inspection. In M. Aliofkhaezai (Ed.), *Developments in corrosion protection*. InTech. <https://doi.org/10.5772/57209>
- Choi, P., Poudyal, L., Rouzmehr, F., & Won, M. (2020). Spalling in continuously reinforced concrete pavement in Texas. *Transportation Research Record: Journal of the Transportation Research Board*, 2674(11), 731–740. <https://doi.org/10.1177/0361198120948509>
- Dorafshan, S., Thomas, R. J., & Maguire, M. (2018). Fatigue crack detection using unmanned aerial systems in fracture critical inspection of steel bridges. *Journal of Bridge Engineering*, 23(10), 04018078. [https://doi.org/10.1061/\(ASCE\)BE.1943-5592.0001291](https://doi.org/10.1061/(ASCE)BE.1943-5592.0001291)
- Fondevik, S. K., Stahl, A., Transeth, A. A., & Knudsen, O. O. (2020). *Image segmentation of corrosion damages in industrial inspections* [Paper presentation]. 2020 IEEE 32nd International Conference on Tools with Artificial Intelligence (ICTAI) (pp. 787–792). <https://doi.org/10.1109/ICTAI50040.2020.00125>
- Forkan, A. R. M., Kang, Y.-B., Jayaraman, P. P., Liao, K., Kaul, R., Morgan, G., Ranjan, R., & Sinha, S. (2022). CorrDetector: A framework for structural corrosion detection from drone images using ensemble deep learning. *Expert Systems with Applications*, 193, 116461. <https://doi.org/10.1016/j.eswa.2021.116461>
- Itzhak, D., Dinstein, I., & Zilberberg, T. (1981). Pitting corrosion evaluation by computer image processing. *Corrosion Science*, 21(1), 17–22. [https://doi.org/10.1016/0010-938X\(81\)90059-7](https://doi.org/10.1016/0010-938X(81)90059-7)
- Jahanshahi, M. R., Kelly, J. S., Masri, S. F., & Sukhatme, G. S. (2009). A survey and evaluation of promising approaches for automatic image-based defect detection of bridge structures. *Structure and Infrastructure Engineering*, 5(6), 455–486. <https://doi.org/10.1080/15732470801945930>
- Javaherdashti, R. (2000). How corrosion affects industry and life. *Anti-Corrosion Methods and Materials*, 47(1), 30–34. <https://doi.org/10.1108/00035590010310003>
- Jeong, E., Seo, J., & Wacker, J. (2020). Literature review and technical survey on bridge inspection using unmanned aerial vehicles. *Journal of Performance of Constructed Facilities*, 34(6), 04020113. [https://doi.org/10.1061/\(ASCE\)CF.1943-5509.0001519](https://doi.org/10.1061/(ASCE)CF.1943-5509.0001519)
- Kesteloo, H. (2019, October 1). Skydio's obstacle avoidance and self-flying capability explained. *DroneDJ*. Retrieved June 2023, from <https://dronedj.com/2019/10/01/skydios-obstacle-avoidance-self-flying-capability/>
- Khayatad, M., De Pue, L., & De Waele, W. (2020). Detection of corrosion on steel structures using automated image processing. *Developments in the Built Environment*, 3, 100022. <https://doi.org/10.1016/j.dibe.2020.100022>
- Khayatad, M., Honhon, M., & De Waele, W. (2022). Detection of corrosion on steel structures using an artificial neural network. *Structure and Infrastructure Engineering*, 19(12), 1860–1871. <https://doi.org/10.1080/15732479.2022.2069272>
- Lin, J. J., Ibrahim, A., Sarwade, S., & Golparvar-Fard, M. (2021). Bridge inspection with aerial robots: Automating the entire pipeline of visual data capture, 3D mapping, defect detection, analysis, and reporting. *Journal of Computing in Civil Engineering*, 35(2), 04020064. [https://doi.org/10.1061/\(ASCE\)CP.1943-5487.0000954](https://doi.org/10.1061/(ASCE)CP.1943-5487.0000954)
- Lin, W., & Yoda, T. (2017). *Bridge engineering: Classifications, Design Loading, and Analysis Methods*. Oxford: Elsevier Butterworth-Heinemann. <https://doi.org/10.1016/C2015-0-00602-5>
- LTBP InfoBridge (US Department of Transportation Federal Highway Administration). (n.d.). [Dataset]. <https://infobridge.fhwa.dot.gov/data>
- Na, S., Xumin, L., & Yong, G. (2010). *Research on k-means clustering algorithm: An improved k-means clustering algorithm* [Paper presentation]. 2010 Third International Symposium on Intelligent Information Technology and Security Informatics, 63–67. <https://doi.org/10.1109/IITSI.2010.74>
- Nash, W., Zheng, L., & Biribilis, N. (2022). Deep learning corrosion detection with confidence. *Npj Materials Degradation*, 6(1), 26. <https://doi.org/10.1038/s41529-022-00232-6>
- Petricca, L., Moss, T., Figueroa, G., & Broen, S. (2016). *Corrosion detection using A.I: A comparison of standard computer vision techniques and deep learning model* [Paper presentation]. Computer Science & Information Technology (CS & IT), 91–99. <https://doi.org/10.5121/csit.2016.60608>
- Rahman, A., Wu, Z. Y., & Kalfarisi, R. (2021). Semantic deep learning integrated with RGB feature-based rule optimization for facility

- surface corrosion detection and evaluation. *Journal of Computing in Civil Engineering*, 35(6), 04021018. [https://doi.org/10.1061/\(ASCE\)CP.1943-5487.0000982](https://doi.org/10.1061/(ASCE)CP.1943-5487.0000982)
- Reynolds, D. (2009). Gaussian mixture models. In: S. Z. Li & A. Jain (Eds), *Encyclopedia of Biometrics.*, 741, 659–663. Boston, MA: Springer. [https://doi.org/10.1007/978-0-387-73003-5\\_196](https://doi.org/10.1007/978-0-387-73003-5_196)
- Sekachev, B., Manovich, N., Zhiltsov, M., Zhavoronkov, A., Kalinin, D., Hoff, B., TOSmanov, Kruchinin, D., Zankevich, A., DmitriySidnev, Markelov, M., Chenuet, M., A-Andre., Melnikov, A., Kim, J., Ilouz, L., Glazov, N., Truong, T. (2020). Johannes222 opencv/cvat: V1.1.0 (Version v1.1.0) [Computer software]. Zenodo. <https://doi.org/10.5281/zenodo.4009388>
- Son, H., Hwang, N., Kim, C., & Kim, C. (2014). Rapid and automated determination of rusted surface areas of a steel bridge for robotic maintenance systems. *Automation in Construction*, 42, 13–24. <https://doi.org/10.1016/j.autcon.2014.02.016>
- Sookpong, S., Phimsiri, S., Tosawadi, T., Choppradit, P., Suttichaya, V., Utintu, C., & Thamwiwatthana, E. (2023). *Comparison of corrosion segmentation techniques on oil and gas offshore critical assets* [Paper presentation]. 2023 20th International Conference on Electrical Engineering/Electronics, Computer, Telecommunications and Information Technology (ECTI-CON), 1–5. <https://doi.org/10.1109/ECTI-CON58255.2023.10153134>
- Tien, I., & Herndon, H. (2023). *UAS-assisted inspection of bridges for corrosion effects* (Final Report FHWA-GA-23-2201). Georgia Department of Transportation, Office of Performance-Based Management and Research. [https://rosap.nrl.bts.gov/view/dot/72841/dot\\_72841\\_DS1.pdf](https://rosap.nrl.bts.gov/view/dot/72841/dot_72841_DS1.pdf)
- Tomiczek, A. P., Whitley, T. J., Bridge, J. A., & Ifju, P. G. (2019). Bridge inspections with small unmanned aircraft systems: Case studies. *Journal of Bridge Engineering*, 24(4), 05019003. [https://doi.org/10.1061/\(ASCE\)BE.1943-5592.0001376](https://doi.org/10.1061/(ASCE)BE.1943-5592.0001376)
- Wu, B., Xu, C., Dai, X., Wan, A., Zhang, P., Yan, Z., Tomizuka, M., Gonzalez, J., Keutzer, K., & Vajda, P. (2020). *Visual transformers: Token-based image representation and processing for computer vision* (arXiv preprint No. 2006.03677). arXiv. <http://arxiv.org/abs/2006.03677>
- Yao, Y., Yang, Y., Wang, Y., & Zhao, X. (2019). Artificial intelligence-based hull structural plate corrosion damage detection and recognition using convolutional neural network. *Applied Ocean Research*, 90, 101823. <https://doi.org/10.1016/j.apor.2019.05.008>
- Zhou, Q., Ding, S., Feng, Y., Qing, G., & Hu, J. (2022). Corrosion inspection and evaluation of crane metal structure based on UAV vision. *Signal, Image and Video Processing*, 16(6), 1701–1709. <https://doi.org/10.1007/s11760-021-02126-7>