

# Automated Bridge Inspection Image Retrieval Based on Deep Similarity Learning and GPS

Benjamin E. Wogen, A.M.ASCE<sup>1</sup>; Jongseong Choi, Ph.D.<sup>2</sup>; Xin Zhang<sup>3</sup>; Xiaoyu Liu, Ph.D.<sup>4</sup>; Lissette Iturburu, S.M.ASCE<sup>5</sup>; and Shirley J. Dyke, Ph.D., A.M.ASCE<sup>6</sup>

**Abstract:** The inspection of highway bridge structures in the US is a task critical to the national transportation system. Inspection images contain abundant visual information that can be exploited to streamline bridge assessment and management tasks. However, historical inspection images often go unused in subsequent assessments because they tend to be disorganized and unlabeled. Further, due to the lack of global positioning system (GPS) metadata and visual ambiguity, it is often difficult for other inspectors to identify the location on the bridge where past images were taken. Although many approaches are being considered toward fully automated or semiautomated methods for bridge inspection, there are research opportunities to develop practical tools for inspectors to make use of those images already in a database. In this study, a deep learning-based image similarity technique is developed and combined with image geolocation data to localize and retrieve historical inspection images based on a current query image. A Siamese convolutional neural network (SCNN) is trained and validated on a gathered data set of over 1,000 real world bridge deck images collected by the Indiana Department of Transportation. A composite similarity (CS) metric is created for effective image ranking, and the overall method is validated on a subset of eight bridge's images. The results show promise for implementation into existing databases and for other similar structural inspections, showing up to an 11-fold improvement in successful image retrieval when compared with random image selection. DOI: [10.1061/JSENDH.STENG-12639](https://doi.org/10.1061/JSENDH.STENG-12639). © 2023 American Society of Civil Engineers.

## Introduction

The US inventory of highway bridge structures is a critical part of its national transportation system. According to the 2021 ASCE annual *Infrastructure Report Card*, approximately 7.5% of the nearly 620,000 bridges in the National Bridge Inventory (NBI) are rated as being in poor condition or structurally deficient (ASCE 2021). Their report on the nation's bridges concluded that the US must drastically increase investment to keep up with deterioration, and a systematic program for preservation is needed that focuses on preventative maintenance (ASCE 2021).

Routine structural inspections, conducted at least every 24 months, are an important part of bridge preservation and provide

the data necessary for asset managers to make timely maintenance decisions (Abdallah et al. 2022). These inspections are generally conducted by visual observation. As a result, condition assessments are largely subjective and result in high variability (Phares et al. 2004; Graybeal et al. 2002). The subjectivity of bridge inspections has motivated a large body of research that aims to develop semi-autonomous or fully autonomous inspection procedures (Dorafshan and Maguire 2018). However, at the present, procedures for routine bridge inspections require that humans perform all inspection observations manually (FHWA 2022). Thus, formal adoption of these capabilities may take some time.

Another opportunity is to empower the human inspector through automation. There is a myriad of data already being stored in asset bridge management databases, and enabling inspectors to capitalize on this data is expected to increase efficiency and effectiveness. To this end, we develop a novel tool for bridge inspectors that enables the rapid recall of past bridge images to support their current decision-making. To support this, our method leverages deep machine learning and global positioning system (GPS) coordinate information to compute a composite similarity (CS) metric between pairs of images, which are then ranked and returned to the inspector. The methodology is validated using a data set of more than 1,000 bridge deck images collected through routine inspections from the Indiana Department of Transportation (INDOT), and the results show promise for application into existing databases.

## Background

Current research aims to develop technologies using advanced computing and artificial intelligence to improve the efficiency and safety of bridge inspections. Toward this vision, researchers are developing systems to gather data autonomously or remotely using robotic agents (Phillips and Narasimhan 2019; Sutter et al. 2018;

<sup>1</sup>Graduate Research Assistant, Lyles School of Civil Engineering, Purdue Univ., West Lafayette, IN 47907. ORCID: <https://orcid.org/0000-0001-5103-0398>. Email: [bwogen@alumni.purdue.edu](mailto:bwogen@alumni.purdue.edu)

<sup>2</sup>Assistant Professor, Dept. of Mechanical Engineering, SUNY Korea, Incheon, South Korea; Dept. of Mechanical Engineering, Stony Brook Univ., Stony Brook, NY (corresponding author). ORCID: <https://orcid.org/0000-0002-6138-8809>. Email: [jongseong.choi@sunykorea.ac.kr](mailto:jongseong.choi@sunykorea.ac.kr)

<sup>3</sup>Ph.D. Candidate, Lyles School of Civil Engineering, Purdue Univ., West Lafayette, IN 47907. ORCID: <https://orcid.org/0000-0001-7615-8013>. Email: [zhan3794@purdue.edu](mailto:zhan3794@purdue.edu)

<sup>4</sup>Postdoctoral Researcher, School of Mechanical Engineering, Purdue Univ., West Lafayette, IN 47907. Email: [liu1787@purdue.edu](mailto:liu1787@purdue.edu)

<sup>5</sup>Ph.D. Candidate, Lyles School of Civil Engineering, Purdue Univ., West Lafayette, IN 47907. Email: [liturbur@purdue.edu](mailto:liturbur@purdue.edu)

<sup>6</sup>Professor, School of Mechanical Engineering, Purdue Univ., West Lafayette, IN 47907; Professor, Lyles School of Civil Engineering, Purdue Univ., West Lafayette, IN 47907. ORCID: <https://orcid.org/0000-0003-3697-992X>. Email: [sdyke@purdue.edu](mailto:sdyke@purdue.edu)

Note. This manuscript was submitted on March 30, 2023; approved on October 9, 2023; published online on December 27, 2023. Discussion period open until May 27, 2024; separate discussions must be submitted for individual papers. This paper is part of the *Journal of Structural Engineering*, © ASCE, ISSN 0733-9445.

Lim et al. 2014) and unmanned aerial vehicles (UAVs) (Montes et al. 2022; Seo et al. 2022; Perry et al. 2020; Kim et al. 2018; Lovelace and Wells 2018). Researchers have also proposed a phased plan for DOTs to transition to one such fully automated inspection system (Lin et al. 2021). These proposed methods involve costly vehicles affixed with equipment such as high-end navigation sensors, infrared cameras, and laser scanners. The inspection robot developed by Phillips and Narasimham (2019), for example, utilized an unmanned ground vehicle (UGV) affixed with a light lidar scanner, an inertial measurement unit, a GPS unit, and an onboard computer.

Researchers are also attempting to automatically quantify and assess structural damage by implementing machine learning algorithms to process images. These computer vision-based approaches have been explored for virtually all elements of a bridge, including the detection of concrete cracks (Deng et al. 2022; Kao et al. 2023; Nguyen et al. 2023), connection bolt issues (Li et al. 2022; Jiang et al. 2023), steel paint condition (Alayub et al. 2022), joint damage (Gagliardi et al. 2022), and for determining substructure vulnerability (Zhang et al. 2022b).

From an implementation perspective, many of these technologies have real or perceived challenges including higher equipment costs, increased inspector training requirements, larger data storage requirements, and reduced driver safety around robots. Although some techniques such as UAV-based inspections are already being utilized for special inspections, scaling these techniques to the entire NBI will be difficult (Lovelace and Wells 2018). Additionally, the general public may be slow to approve of inspection automation where life safety is of primary concern (Shariff et al. 2017). These challenges must be overcome before DOTs are able to implement them as a part of their processes. A predecessor to autonomous data collection and interpretation may be to provide inspectors with automated systems in a way that they still make all decisions themselves (Jiang et al. 2004).

Automating and improving image organization is one way for inspectors to get more information out of their existing bridge databases. Inspection images contain information on the temporal condition of a structure, but they often go unused in subsequent inspections. Tracing the history of structural defects over time through past images, inspectors can analyze deterioration rates, which may prove critical for timely preventative maintenance actions (Bianchi et al. 2022). Although sifting through these images is a useful inspection exercise, there may be hundreds of images stored for a given bridge, and the images are often disorganized and unlabeled (Zhang et al. 2022a). Further, due to visual ambiguity, it is often difficult for the next inspector to identify the location on the bridge where a past image was taken from just the image and its labels, if any labels are present (Chun et al. 2022; Yamane et al. 2023).

We develop a solution to aid human inspectors that enables the automatic identification of past inspection images corresponding to the same location and scene on a bridge. To assess the similarity of image pairs based only on information already in the databases, the authors leverage three extractable pieces of information, including (1) image GPS coordinates; (2) classification of the image in terms of its structural element category; and (3) visual similarity of the image to the current image of interest. The first item is the geolocation of the images; images taken close to each other in space are more likely to contain the same feature of the bridge. However, many images include no geolocation data at all, whereas others include inaccurate data.

The second and third pieces of information are extracted from the images themselves using computer vision techniques. Deep convolutional neural network (CNN)-based classifiers have proven to be successful at sorting images rapidly and accurately into predefined categories (Krizhevsky et al. 2017). Within structural

engineering, researchers have leveraged this capability for damage detection and structural element recognition in a variety of applications (Zhang et al. 2016; Cha et al. 2017; Spencer et al. 2019; Yeum et al. 2019; Zhang et al. 2022b; Yu and Nishio 2022). Using this technique, a large image set for a given bridge can be quickly and reliably sorted into elemental categories. Sorting can serve to narrow a pool of relevant prior images that may be similar to one another.

The final piece of information available is the assessment of image similarity via deep machine learning. Despite variation over time in the images of certain bridge features due to changing lane markings, construction activities, seasonal weather, and deterioration, images of similar locations and defects on the structure may share certain common visual features that can be learned by a Siamese CNN (SCNN) and compared using available similarity metrics. Variations of this technique are used for applications in facial recognition (Cao et al. 2013; Schroff et al. 2015; Apple 2021) and analyzing medical imagery (El-Naqa et al. 2004; Yang et al. 2010; Ktena et al. 2017). More recently, researchers within civil engineering have applied SCNNs to assist with postreconnaissance analyses (Choi et al. 2022) and for detecting railroad track damage (Yang et al. 2023).

## Methodology

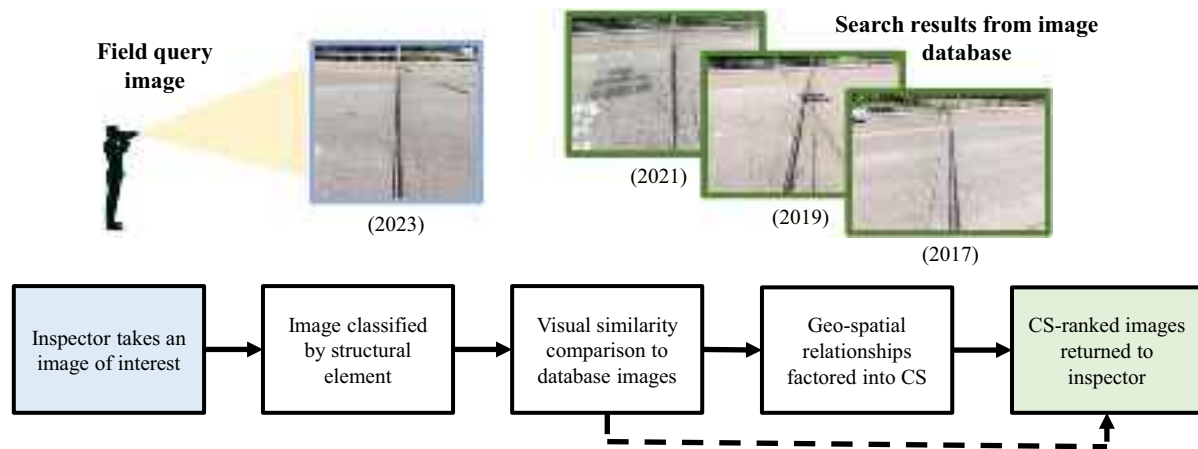
### Overview

The developed method leverages previous similarity-based applications (Choi et al. 2022) in other fields by fusing image similarity with geolocation information. Neither image similarity or GPS coordinate relationships alone provide sufficient information for this application, and thus we develop a method that is able to consider both. The overall methodology is illustrated in Fig. 1.

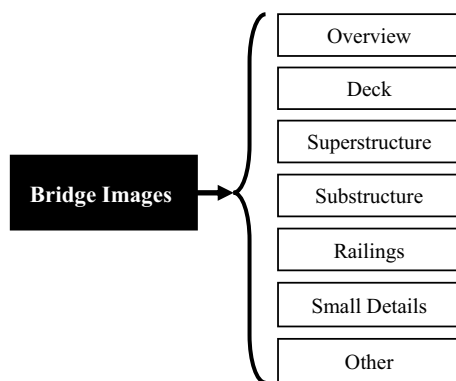
First, a bridge inspector takes a photograph of an element of interest on a given bridge, hereafter referred to as the query image. Second, the query image is classified based on the structural element that it contains (e.g., deck, superstructure, and substructure). Third, using a trained SCNN, the query image is reduced in dimensionality and mapped to an invariant embedding vector. The network is trained such that images with similar learned features are mapped to spatially similar vectors, whereas images with different learned features are mapped to spatially different vectors. The query image embedding is then compared with the embeddings of all other images in a given bridge's database within its element category, returning a ranked list of the most similar images to the inspector. If geolocation data are available for a given pair of images, the scores are adjusted using a novel CS scoring metric that combines image similarity and with geospatial data. Finally, the most relevant prior images are returned to the inspector based on the ranked CS values. In this way, an inspector can capture an image and be provided with past images of the same location and damage in real-time.

### Structural Element Classification

Once the query image is captured or selected, it must be classified. Recent research used CNN-based image classification to automatically sort bridge inspection images per the elements defined in the *AASHTO Manual for Bridge Element Inspection* (MBEI), which is the basis for current element-level bridge inspection in the US (Zhang et al. 2022a; AASHTO 2013). Zhang et al. (2022a) developed the classification schema shown in Fig. 2 and validated their method using a data set of more than 11,000 images collected by inspectors, with an average accuracy of 94%.



**Fig. 1.** Workflow for the image localization and retrieval method. (Images courtesy of Indiana Department of Transportation.)



**Fig. 2.** Classification schema applied in the developed method. (Reprinted from Zhang et al. 2022a.)

This previously developed classifier was applied in the current image retrieval task. Classification is necessary to narrow the search pool of images for retrieval, creating a higher likelihood of finding query matches quickly. Then, in this workflow, a separate similarity-step SCNN must be trained for each element category. For example, the deck category will have its own independently trained image-search capability, as will the substructure category, and so on.

### Image Similarity Training

With trained elemental classifiers, the next step in the workflow is the comparison of the query image with each of the prior images from the given bridge in the same element category. This step is not suitable for CNN-based classification because the locations and angles of the deck images do not fit into any predefined categories. Instead, an image retrieval capability is developed using a training method established by Hadsell et al. (2006). This technique learns the parameters  $W$  for some function  $G_W$ , in this case a deep CNN, to reduce the dimensionality of images to a lower dimension vector by a learned invariant mapping (also referred to as an embedding).

For a given set  $\mathcal{L}$  of  $n$  images corresponding to a single bridge element, Image 1 is represented in terms of its pixel components as a high-dimensional vector,  $\vec{X}_1$ . This image may have a subset of other images in  $\mathcal{L}$  that are similar to it by some defined measure. For the large size of the vector  $\vec{X}_1$ , this measure would be difficult to

compute. However, by an appropriate learned function  $G_W$ ,  $\vec{X}_1$  may be mapped to a smaller dimension vector,  $\vec{x}_1$ . If each  $\vec{X}_i \in \mathcal{L}$  for  $i = 1, 2, 3, \dots, n$  is mapped to a smaller  $\vec{x}_i$  in an invariant way, Image 1 can then be compared with every other image using the Euclidean distance between their reduced-dimension vectors, as represented for a given pair  $(\vec{X}_1, \vec{X}_2)$  by  $D_W$  in Eq. (1) (Hadsell et al. 2006) as follows:

$$D_W(\vec{X}_1, \vec{X}_2) = \|\vec{x}_1 - \vec{x}_2\|_2 = \|G_W(\vec{X}_1) - G_W(\vec{X}_2)\|_2 \quad (1)$$

This distance is then used to calculate a loss  $L$  for the pair of images in the contrastive loss function provided by Eq. (2) (Hadsell et al. 2006) as follows:

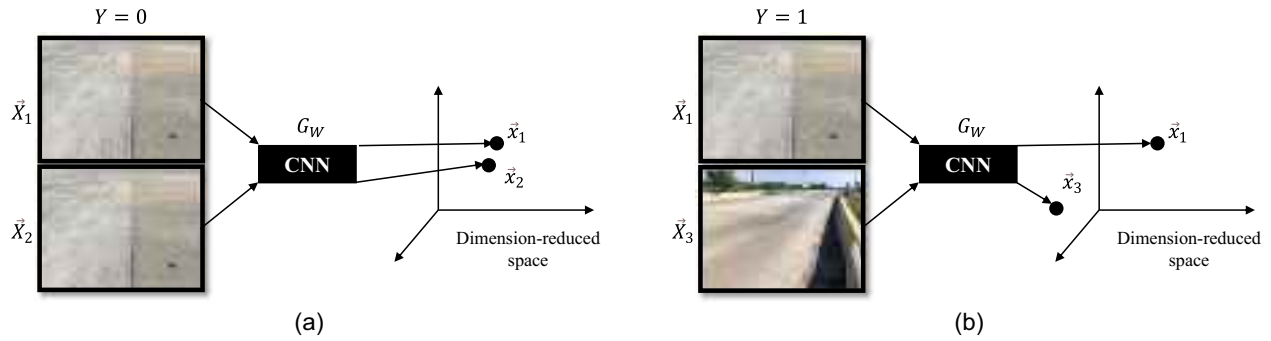
$$L = (1 - Y) \frac{1}{2} D_W(\vec{X}_1, \vec{X}_2)^2 + Y \frac{1}{2} \{\max(0, m - D_W(\vec{X}_1, \vec{X}_2))\}^2 \quad (2)$$

where  $Y$  = labeled relationship between Images 1 and 2, and thus the loss function is different based on a pair's label. For image pairs defined as similar,  $Y = 0$ , and the loss is determined only by the first term. In this case, if the embedding distance  $D_W$  is small, then the loss is small, and the function  $G_W$  has mapped a similar pair close by. On the other hand, if  $D_W$  is large, the loss is large, and the function  $G_W$  has mapped a similar pair far apart.

In contrast, image pairs defined as dissimilar are labeled with  $Y = 1$ , and the loss is determined only by the second term. In this second term, the parameter  $m$  is defined as the margin, a predefined parameter that is always greater than zero. Hadsell et al. (2006) described the margin as a limit to how dissimilar pairs can be while still contributing to the loss. In this case, if  $D_W$  is large, then the loss is small, and  $G_W$  has mapped a dissimilar pair far apart. However, increasing  $D_W$  values will have an impact only up to the margin. In this way, the margin exists to eliminate cases with abnormally high Euclidean distance measures (Choi et al. 2022). On the other hand, if  $D_W$  is small, then the loss is large, and  $G_W$  has mapped a dissimilar pair close by. A simplified illustration of the embedding process using a three-dimensional output is provided in Fig. 3 because it is easy to visualize; however, any number of dimensions may be used for the embedding. In Fig. 3(a), the pair of Images 1 and 2 represents a similarly defined pair and is labeled with  $Y = 0$ , and in Fig. 3(b), Images 1 and 3 represent a dissimilarly defined pair and are labeled with  $Y = 1$ .

The training occurs using a SCNN architecture in batches of paired images, where the total loss for a given training batch is





**Fig. 3.** Representation of the embedding process for (a) a similarly defined pair; and (b) a dissimilarly defined pair. (Images courtesy of Indiana Department of Transportation.)

the sum of the losses for each image pair. In the learning process for a single pair, both images are processed in parallel by two different but identical (Siamese) networks ( $G_W$ ), and the loss for the pair ( $L$ ) is calculated. During each training epoch, the parameters  $W$  in the twin networks are adjusted equally such as to minimize the loss function using gradient descent (Hadsell et al. 2006). Thus, the outcome of training is a single network that is trained to reduce the dimensionality of images in such a way that the embeddings are similar or different based on the labeling definitions.

### GPS-Refined Composite Similarity

Once the network is trained, it can be used to map any new image in the corresponding element class. For the image vector pair given by  $(\vec{X}_1, \vec{X}_2)$ , the embedded distance is used to calculate a normalized similarity score (SS). Because the results will never be perfect, the intention is to retrieve the most similar images in the form of a ranked list. To rank the prior images, a scoring system is needed. Choi et al. (2022) showed success with the similarity score (SS) defined as follows:

$$SS(\vec{X}_1, \vec{X}_2) = \left( \frac{1}{1 + D_W(\vec{X}_1, \vec{X}_2)} \right) \quad (3)$$

This function serves to normalize the Euclidian distance,  $D_W$ , to a value between zero and one. When the embedding distance between two image embeddings is large, the SS approaches zero. Alternatively, when the embedding distance between the two images is small and the images are learned to be dissimilar, the SS approaches one.

For this application, the SS concept is improved by considering the added benefit that image geolocation may have in the ranking process. A CS score is thus developed that fuses the coordinate distances between images into the same domain as the SS. This equation may also be optimized based on the quality and quantity of the GPS metadata in the given bridge database by adjusting its parameters. The CS score for the pair of Images 1 and 2 can be calculated using Eq. (4) and is the superposition of two weighted components

$$CS(\vec{X}_1, \vec{X}_2) = (1 - \beta) \left( \frac{1}{1 + D_W(\vec{X}_1, \vec{X}_2)} \right) + \beta \times \max \left( p_{\min}, \left( 1 - \frac{d(\text{GPS}_1, \text{GPS}_2)}{d_{th}} \right) \right) \geq 0 \quad (4)$$

The first term is taken from Eq. (3) and reflects the output of the trained similarity model in a score ranging from zero to one.

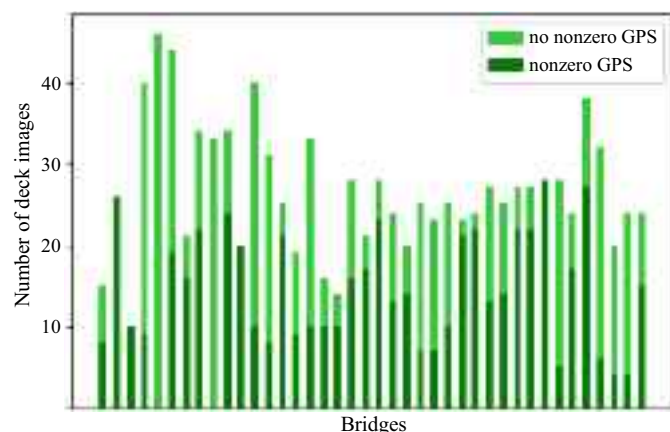
The second term is the geolocation term and assumes that similar images will also have been taken at similar points in latitude and longitude coordinate space. The function  $d$  is the Euclidian distance (m) between the coordinates  $\text{GPS}_1$  and  $\text{GPS}_2$  of Images 1 and 2, respectively, and thus can be any positive value. The constant  $d_{th}$  is chosen based on an analysis of the data set and represents a threshold distance for whether to penalize or reward the CS score. For image pairs where  $d$  is small, the images are taken sufficiently close together, and thus the second term is positive and approaches one. As  $d$  increases but remains smaller than  $d_{th}$ , the images are less likely to be depicting the same scene, and this term of the equation approaches zero. When  $d > d_{th}$ , the images are unlikely to show similar locations, and the score may become negative. The minimum geolocation score component can be defined by the database manager as  $p_{\min}$ .

The weight  $\beta$  in Eq. (4) represents the weight given to the geolocation term relative to the visual similarity term and serves to keep the total composite score normalized between zero and one. This term can be optimized for the quality and quantity of the images and their coordinates and may evolve over time with better GPS accuracy. For example, if a data set has geolocation data that is highly inaccurate, it may make sense to have  $\beta = 0.10$ . In this case, the CS score will be determined mostly by the visual similarity from Eq. (3), and the GPS component will be only a slight adjustment. For GPS data to be considered, both images being compared must have coordinates. Thus, for any specific image pair where one or both images lack geolocation data, the term  $\beta$  must be taken as zero, and the score for that pairing is only represented by Eq. (3). In this way, the developed method applies to databases with varying proportions of GPS availability.

## Experimental Validation and Results

### Bridge Deck Data Set Overview

The first step of image classification was previously validated by Zhang et al. (2022a) with an average accuracy of 94%. Thus, this step will not be validated in this work. Rather, a subset of those bridge images was chosen for validating the novel proposed method and was assumed to be already classified. Specifically, images from the deck category were chosen for several reasons. First, deck images are more likely than other categories to have accurate GPS metadata. The geolocation accuracy of mobile-device GPS is significantly reduced if overhead obstructions are present, such as those images in other categories taken under the bridge deck (e.g., the substructure category). Although forest canopies, mountains, and buildings may also impact the accuracy of recorded



**Fig. 4.** Gathered data set containing historical deck images for 40 bridges with varying GPS coordinate availability. Approximately half of the images (52%) have nonzero GPS.

coordinates, the deck category has the highest potential for accurate metadata (Merry and Bettinger 2019; Wing et al. 2005). In addition, bridge decks are generally the fastest deteriorating bridge components and thus should be a main concern when considering inspection advancements (Kong et al. 2022).

A total of 1,068 deck images taken from 40 different bridges were provided by the INDOT. The data stored for these bridges had an average of 27 deck images each, with variations in the existence and reliability of the GPS metadata, as shown in Fig. 4. Altogether around half of the images (52%) had nonzero GPS coordinates. It is important to specify these as nonzero coordinates because many of the images stored in the database were assigned to have zero-valued coordinates, likely due to inspectors processing them through different interfaces and software. The method must remain effective even when no geolocation metadata are available, as is the case for several of the bridges in the data set.

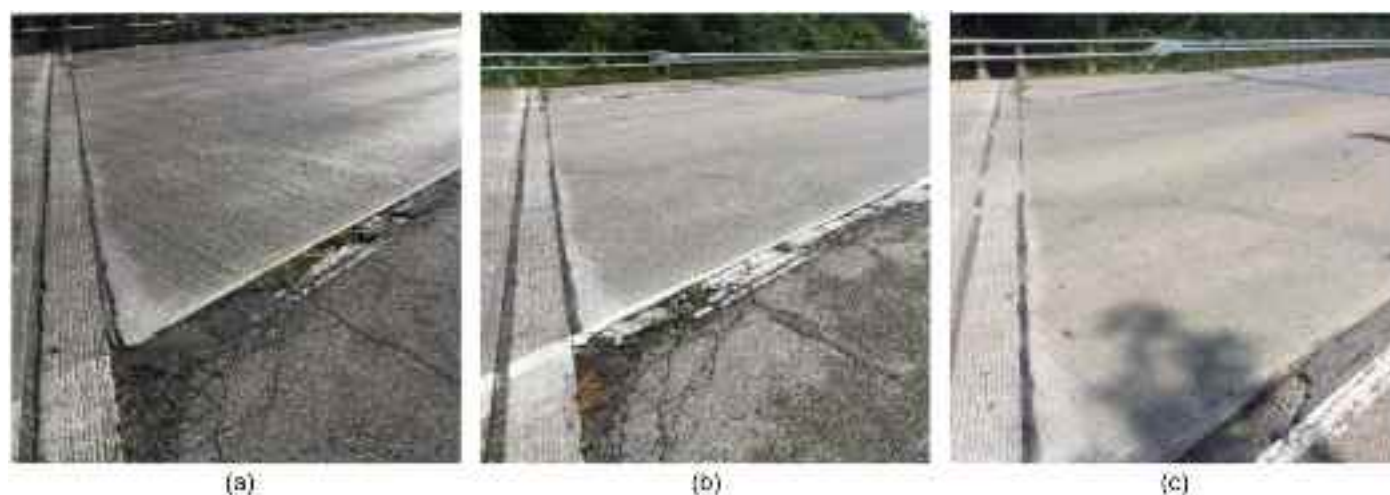
Several examples are shown in Fig. 5, representing a set of similar images. This figure illustrates the potential utility of an automated image localization and retrieval system. In a perfect application of this framework, an inspector in the field may capture

the query photo in Fig. 5(a) and then be returned the images shown in Figs. 5(b and c) corresponding to two previous inspection cycles. These images have variations due to weather, camera type, deck deterioration, and maintenance actions, making the application of similarity a challenging task. Fig. 5 is also representative of the quality of the GPS metadata in the data set. GPS information may be present and reasonably accurate, present but inaccurate, or missing entirely. Because no ground truth exists for the image coordinates for the data set used for training and validation, the ground-truth locations are estimated using Google satellite maps (Google 2023). As collected from the database, only about half of the images contain GPS metadata, and the coordinates that do exist often have high deviations from the estimated actual locations.

### Labeling Definitions and Challenges

Although the SCNN does not measure images in the sense of percent overlap or angle differences, having precise definitions for what constitutes a true match versus a false match serves to keep the labeling consistent across the entire data set. A true match pair was defined as any pair of images captured from the same bridge deck, which capture approximately the same location of the deck from a similar angle, and that share similar visual features. In training, these image pairs were labeled with  $Y = 0$ . A false match pair was defined as any pair of images captured from the same bridge deck, but that capture different locations or features of the deck and otherwise do not meet the requirements for a true match. In training, these image pairs are labeled with  $Y = 1$ . Image pairs taken from different bridge decks are not used in the training or evaluation of the model.

The authors acknowledge that the inspectors likely do not care about the differing perspectives or angles of image pairs so long as they are showing the given feature of interest. However, the authors found these more restrictive definitions necessary for the successful recognition of similar images using the SCNN architecture and for the inclusion of GPS in the composite scoring. Overall, this does not impact the usefulness of the proposed methodology because the inspectors only need one true match per inspection cycle to be able to trace the deterioration of the structure.



**Fig. 5.** Sample timeline of similar deck images taken from the data set for a single bridge, with GPS metadata extracted from the image files and errors calculated from the estimated ground truth: (a) August 2020, no GPS data; (b) August 2018, estimated GPS error = 20 m; and (c) August 2015, estimated GPS error = 9 m. (Images courtesy of Indiana Department of Transportation.)



**Fig. 6.** Examples of labeling for typical bridge images. (Images courtesy of Indiana Department of Transportation.)

Several examples of challenging labeling situations are provided in Fig. 6, including a deck joint [Fig. 6(a)], a deck overview image [Fig. 6(b)], and a close-up wearing surface image [Fig. 6(c)]. Deck joints are some of the most frequently photographed components of a deck and contain many details. These images pose a challenge because joint images taken from different sides of the deck or different sides of the bridge have the potential for high visual similarity. The overview images shown in Fig. 6(b) are also common but may have more variation in their GPS coordinates, even for true match pairs. This variation is because they are captured from zoomed-out perspectives, and so images at far apart coordinates may be labeled as true matches and share similar visual features. Likewise, wearing surface images like those shown in Fig. 6(c) pose a challenge because it is impossible to determine if they were taken at the same location or not.

The authors follow the prescribed labeling definitions while also considering the maximum utility of the similar pairs to the inspectors based on the intention of the images. For example, for the wearing surface images shown in Fig. 6(c), the zoomed-in images are all labeled as similar pairs, even if it is unlikely that they are capturing the same exact location on the deck. This choice was made because these images serve the purpose of measuring the widest bridge cracks and are meant to be representative of the entire wearing

**Table 1.** Summary of the labeled data set as divided into training and testing subsets

Data split	No. of bridges	No. of images	No. of true matches	No. of false matches
Training	32 (80%)	873	902	11,543
Testing	8 (20%)	195	161	2,353
Total	40	1,068	1,063	13,896

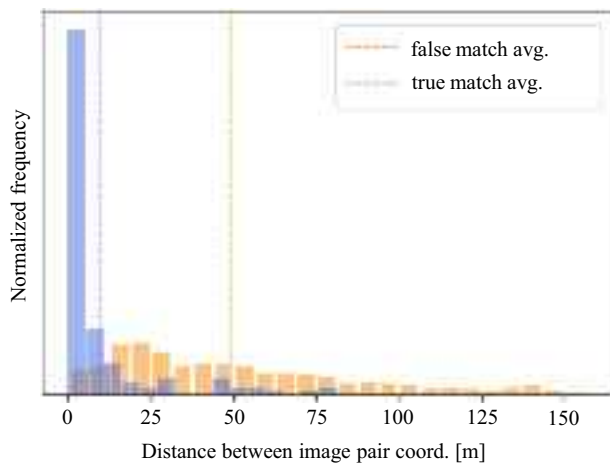
surface of the structure at that point in time. Thus, these pairs are labeled as true matches so that inspectors can see the overall wearing surface condition in any given year as a result of their image search.

Overall, for the gathered 40-bridge data set, 14,959 image pairings were manually labeled as prescribed in this section. The labeled data set is summarized in Table 1, including the distribution of the training and testing sets described in the next section.

### GPS Ground-Truth Estimation

As mentioned previously, to utilize the CS score developed in this paper, ground-truth coordinates were estimated using satellite





**Fig. 7.** Distribution of satellite-estimated ground-truth distances for all image pairs in the validation data set.

imagery from Google Maps (Google 2023). In this way, the point where the camera is located was estimated for each of the 195 testing images. Although not perfect, the authors estimate accuracy to be within 2 m for the majority of images, resulting in the best data available for ground-truth geolocation data. Fig. 7 shows the distributions for both the true matches and false matches of the Euclidian distance between each testing pair of images.

As predicted, true match pairs were generally taken close together in space, whereas false match pairs were taken farther apart and with more variation. The dashed lines in Fig. 7 represent the mean values of 9 and 50 m between the image coordinates of the true match and false match pairs, respectively.

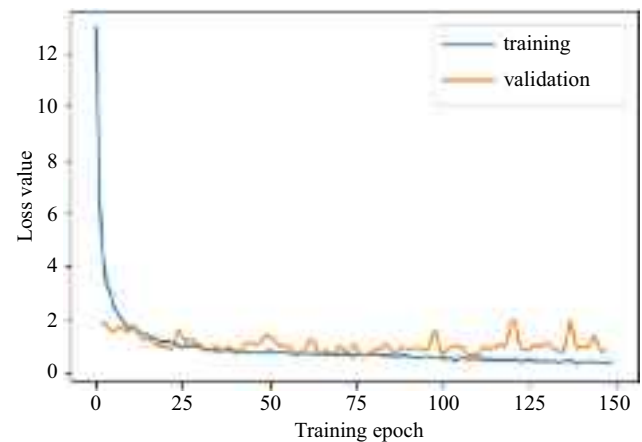
### Similarity Training Setup

The data set was split randomly into a training data set and a testing data set at a 4:1 ratio, as summarized in Table 1. A total of 873 images belonging to 32 bridges were used to train the model, and 195 images belonging to eight bridges were used in validation. Although this may seem like a small data set for a deep learning task, the training and validation occur over pairs of images, not the images themselves. Thus, among the 873 training images, there were more than 12,445 unique same-bridge image pairs (902 true matches and 11,543 false matches).

For training, all images were cropped to the maximum square size from the center point of the original image and then downsized to  $100 \times 100 \times 3$ . All three RGB color channels were preserved because the colors may contain valuable information for details such as lane markings. Several image sizes were tested, and downsizing the images was found to have little effect on the results.

In addition, data augmentation was applied in the form of random rotations of  $\pm 10^\circ$ , contrast changes of  $\pm 10\%$ , brightness changes of  $\pm 10\%$ , and saturation changes of  $\pm 10\%$  to avoid overfitting (Choi et al. 2022). Zoom augmentations were not implemented such as to preserve the labeling intentions and to not accidentally omit similar deck features.

The network architecture chosen as the baseline model was MobileNetV3 (Howard et al. 2019). A batch size of eight image pairs is used, and the learning rate was set to  $5 \times 10^{-4}$ . A NVIDIA (Santa Clara, California) TX TITAN X graphics processing unit (GPU) was utilized to perform the training, and 150 training epochs were chosen empirically. The margin,  $m$ , was set to 2.0, and the



**Fig. 8.** Sample of the loss history for 150 training epochs.

output embedding size was chosen to be 10 based on trial and error. The resulting loss function given by Eq. (2) was calculated for both the training and validation set at each epoch, as shown in Fig. 8.

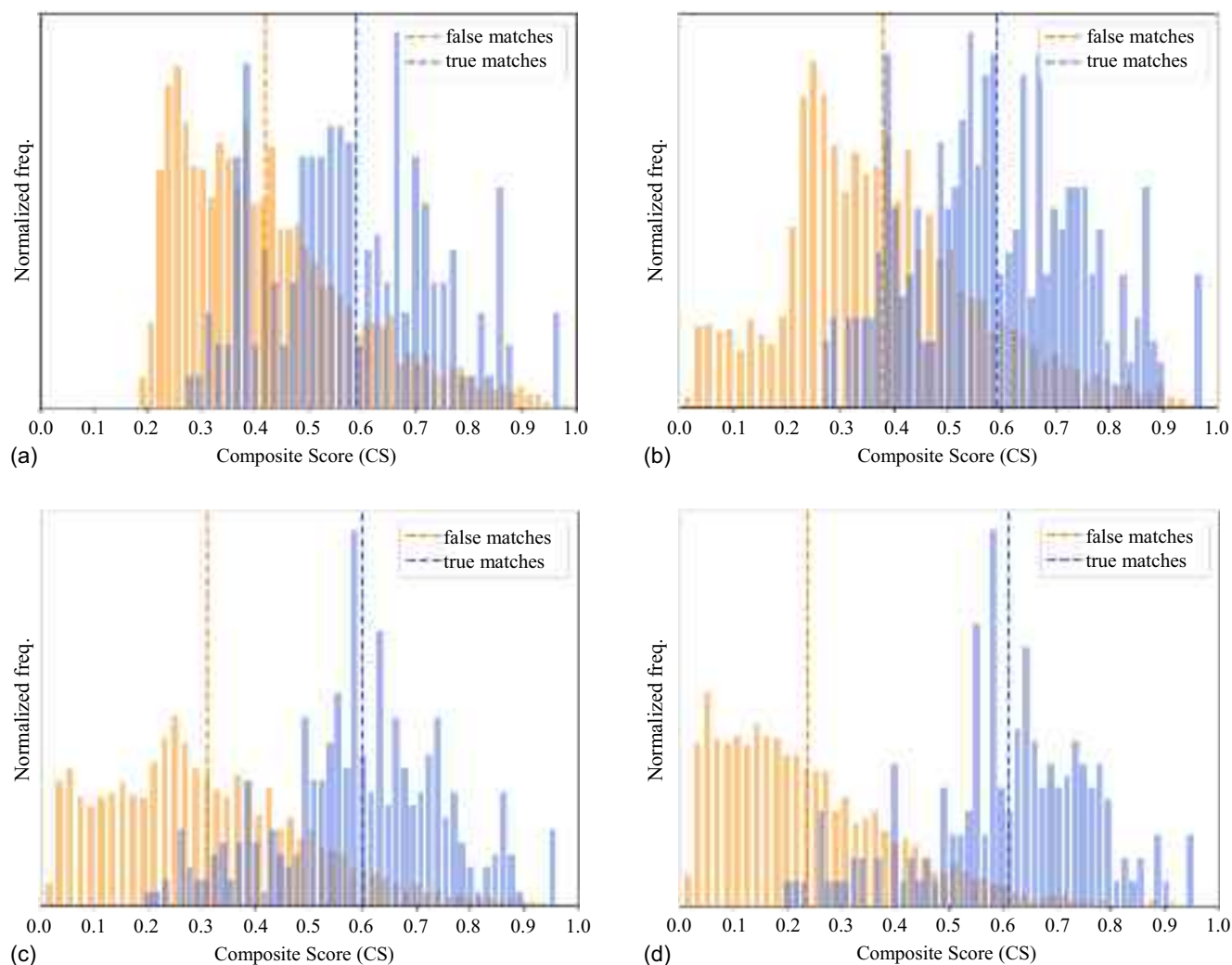
Overall, the training was successful, and the loss appeared to converge to a minimum value after 150 epochs. The high noise and small loss improvement exhibited in the validation loss was due to the small size of the batches, which was necessary due to computing limitations. This behavior is acceptable because the similarity learning is validated in the next section and shows that the network is successful at distinguishing between true and false match pairs in the testing data set. The training requires approximately 2 h on average for the given hardware and data processing specified in this section.

### Composite Similarity Scoring and Ranking

Once the similarity model was trained, it was evaluated by computing CS scores [Eq. (4)] for all of the 2,514 possible image pairings in the testing data set and comparing the CS distributions for labeled true matches and labeled false matches. Further, for every query image, the position of the highest-ranked true match within a ranked list of same-bridge images was evaluated, referred to as the similarity rank (Choi et al. 2022). Based on the GPS analysis in Fig. 7, a threshold distance  $d_{th} = 20$  m was chosen for Eq. (4) as a reasonable value for this data set because most true match pairs have a distance falling below this value. Further, a maximum penalty of  $p_{min} = -0.40$  and weight of  $\beta = 0.30$  were chosen to illustrate the results.

Using these parameters, the CS score was calculated for each image pair, and the resulting distributions are shown in Fig. 9. The different plots in Fig. 9 represent different percentages of images having geolocation metadata in the testing set, where Fig. 9(a) neglects all GPS, Fig. 9(b) shows sample GPS for 50% of images, Fig. 9(c) shows sample GPS for 75% of images, and Fig. 9(d) considers all GPS. The resulting spread between the average true match CS scores and the average false match CS scores (the distance between the dashed lines) increased from 0.17 in Fig. 9(a) to 0.36 in Fig. 9(d) when the estimated ground-truth GPS data were added to the data set.

The specific images having coordinate information for each comparison were sampled at random. Recall that both images in a combination must have GPS information to have a nonzero  $\beta$  term in Eq. (4). Therefore, the percent of combinations that consider GPS data is equal to the square of the percent of images with GPS, and the number of comparisons made where both images



**Fig. 9.** Similarity score distributions for all pairs in the testing data set, shown for an example where  $\beta = 0.30$ ,  $d_{th} = 20$  m, and  $p_{min} = -0.40$ , for varying rates of GPS availability: (a) 0% GPS images; (b) 50% GPS images; (c) 75% GPS images; and (d) 100% GPS images.

have GPS information increases exponentially as the proportion of images with GPS increases. To visualize the results in Fig. 9, both the true match and false match CS score distributions were mass-normalized due to the much higher quantity of false match pairings.

In a similar way, the performance was evaluated by ranking the CS results for each query image. Each image in the testing data set was used once as the query image, and the CS score was calculated between that image and all the other same-bridge deck images. In the final step of the method from Fig. 1, all prior images are ranked by their CS and returned to the user. One such example of this ranking process is shown in Fig. 10, where a true match is returned as the first ranked image. Because the inspector alone knows which precise feature they are interested in, they ultimately choose which images to analyze based on the ranked list. In this way, the number of images an inspector must look through before finding a true match pair is an important performance metric. The distributions of these resulting similarity rankings of the first true match for each query image are illustrated in the histogram in Fig. 11.

Fig. 11 illustrates that as more of the images in the database had quality GPS information, the true matches were able to be retrieved earlier in the CS-ranked list. The performance of the similarity search is further summarized in Table 2 for several values of

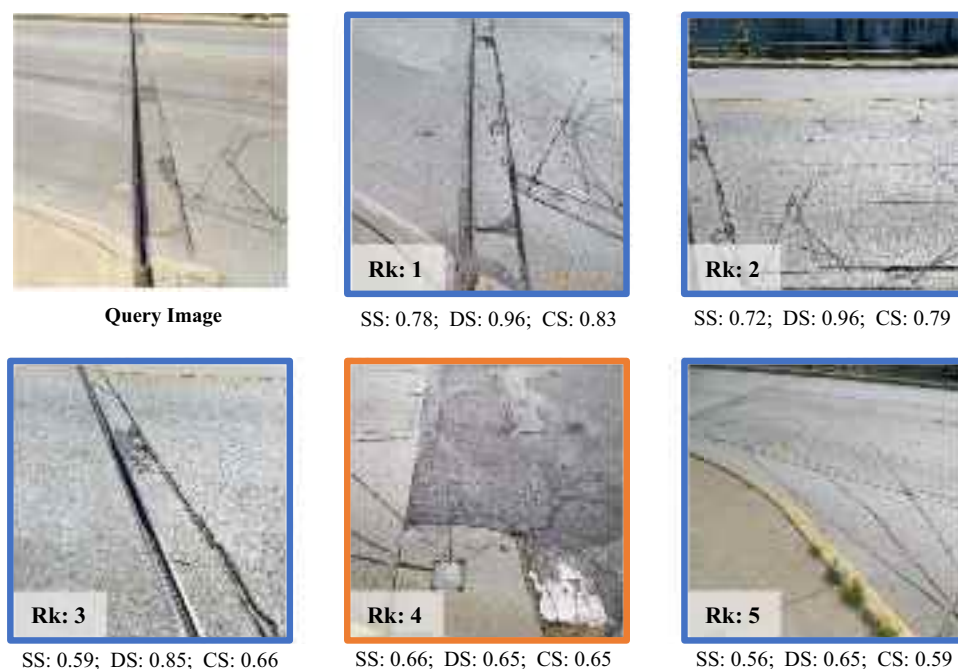
interest. The theoretical percentage of random selection was determined using the average number of images per bridge deck ( $\sim 27$ ), and the average number of true matches per image (approximately one). Images that had no true matches (36 of the 195 in the testing data set) were not considered in this analysis. Table 2 indicates that the proposed method performs favorably to random selection.

## Discussion

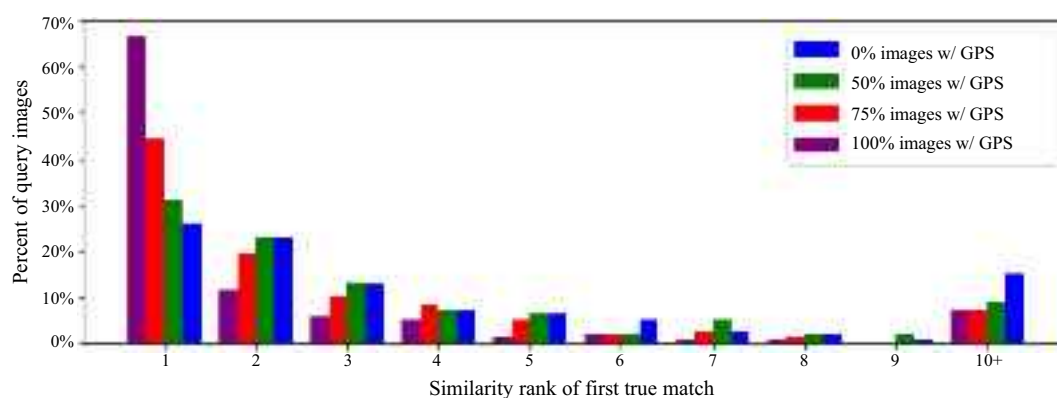
### Overall Performance

The results show that the method performs favorably and has promise for immediate application in bridge management databases. An image search capability designed specifically for structural inspection will enable inspectors to readily make use of historical bridge images in their current procedures. The results of this model when considering no GPS information were worse than previous implementations (Choi et al. 2022), and this result is likely due to higher variability in the true match definitions. Despite true matches being restricted to image pairs showing somewhat similar perspectives of a given feature of the bridge, there was still a lot of variability in





**Fig. 10.** Example results showing the top five retrieved images for a given query image, out of 39 possible matches. True match pairs and false match pairs are indicated. SS, distance score (DS), and CS are shown for each image with respect to the query image. (Images courtesy of Indiana Department of Transportation.)



**Fig. 11.** Similarity rank distributions for all starting query images in the testing data set for the same parameters provided in Fig. 9. ( $\beta = 0.30$ ,  $d_{th} = 20$  m, and  $p_{min} = -0.40$ )

**Table 2.** Similarity ranking success for different proportions of database images having GPS

Images in data set with GPS (%)	Average true match similarity rank	True match ranked first (%)	True match ranked in top 3 (%)	True match ranked in top 5 (%)
0	4.9	26	61	75
50	4.0	31	67	80
75	3.3	44	74	87
100	2.5	66	84	90
Theoretical random selection		4	12	20

those angles and levels of zoom, as well as variability due to the aging and maintenance of the bridge between cycles. Further, among true match pairs, the background scenery varied significantly between inspection cycles due to seasonal changes and road-side developments (e.g., new buildings). Additionally, some error in the results is likely attributable to the high similarity of false

match pairs. Bridge deck features, when compared with buildings, tend to look very similar (i.e., joints) even if they are at different locations of the deck. However, the novel GPS-fusion approach greatly increased the effectiveness of the image queries.

The method can be applied with acceptable performance even when no GPS data are available for the images in the database.

From the results in Table 2, compared with randomly selecting a single past deck image, a user has a six times better chance of selecting a relevant image (true match) using image similarity. If all GPS data are considered and contain minimal error (as was assumed for this analysis) the results are improved, and an inspector has an 11 times greater likelihood of selecting a relevant image (true match) compared with random selection. Moreover, 90% of the time, the inspector will find what they are looking for within the top five CS-ranked images.

Although comparing this method to random selection is a relatively low baseline, it does reflect the current state of practice where an inspector must manually sift through images. This improvement over random selection is likely underestimating the performance in practice when considering the entire method, including the first image classification filter. This behavior is because the ~27 searchable images per bridge are exclusively deck images, whereas a full bridge data set may contain more than 100 disorganized images.

### Composite Scoring Optimization

The results assume that all GPS data are reasonably accurate, at least as much as the ground-truth estimation process would allow.

Although still improving, geolocation technology is still prone to error, especially when obstructions and foliage are present (Merry and Bettinger 2019; Wing et al. 2005). This error is evident in the original GPS for the data set collected, which sometimes has errors on the scale of kilometers. For a given database, the parameters in Eq. (4) may be optimized for the estimated error present. For example, if an inspector knows that the quality of their GPS data is poor, they may elect to set  $\beta = 0$  and ignore it entirely. On the other hand, if an inspector has perfect GPS data, they could elect to set  $\beta = 1$  and consider only GPS coordinates in their searches.

An example of optimal  $\beta$  selection for a data set with artificial error is illustrated in Fig. 12. For this data set, a random error between 0 and 50 m was added to the distance between the ground-truth coordinates for each image pair. Then, the method was again evaluated based on the performance metrics discussed in the results. The spread of the average true match CS scores and false match CS scores are given for varying GPS availability and  $\beta$  values in Fig. 12(a). Similarly, the percentages of all queries returning a true match within the top three rankings are provided in Fig. 12(b). Due to the randomness in which images are assigned to GPS and which GPS comparisons are assigned with higher errors, this process was followed for 1,000 iterations of  $\beta$  ranging from

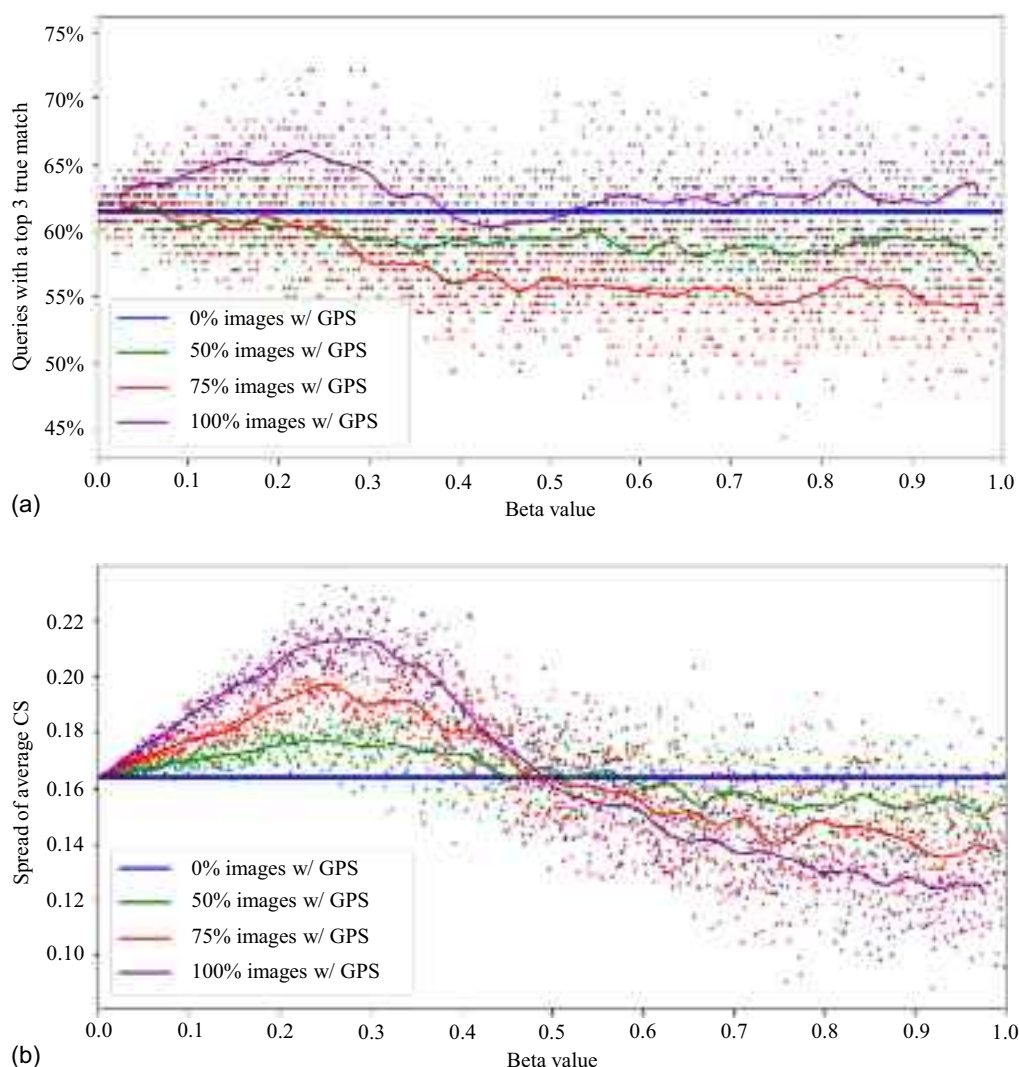


Fig. 12. Optimizing  $\beta$  for a database with moderate GPS error (0–50 m).

zero to one to determine the value that, on average, maximizes the two chosen performance metrics.

The database manager can observe that when  $\beta = 0$ , the metrics converged to those of the 0% GPS data set. This is because when no GPS data were available for one or both images in a pairing,  $\beta$  was always taken to be zero. From the figure, the rank metric in Fig. 12(a) is optimal for values of  $\beta \approx 0.15$  but might vary based on the proportion of GPS data available. Additionally, the spread metric in Fig. 12(b) is optimal for values of  $\beta \approx 0.25$ , regardless of GPS availability. Thus, for the hypothetical error-prone data set created, the value  $\beta \approx 0.20$  should be chosen to optimize overall performance. In this way, inspectors can tailor this method to the quality of their historical image data.

### Practical Considerations

Based on the results, the authors have several recommendations for practicing bridge inspectors that would improve the performance of the method. First, inspectors should take care when capturing and transferring images to preserve all metadata recorded by the camera. Further, if accurate camera bearing information is gathered and observed, it will provide an opportunity to refine the CS method. In addition, standardized and consistent image practices should be used. For deck images, the authors recommend (1) two alignment photos, one from each lengthwise side of the bridge; (2) two photos of each bridge joint and approach slab, one from each widthwise side of the bridge deck; (3) one zoomed-in wearing surface photo representative of the entire deck; and (4) extra images of any notable damage. Following this recommendation, more than a dozen unique images would be taken of a bridge deck during any given inspection. Using image retrieval, a continuous temporal sequence of images could then be observed for each location of the deck, as shown in Fig. 4. Further, with such standardized data, the model training would also likely improve compared with the results presented herein because this model is trained with highly nonuniform image locations and angles.

### Conclusion

Bridges are crucial elements in our national transportation networks, and they are deteriorating rapidly. Bridge asset managers are under pressure to make the current inventory last longer to put off eventual reconstructions. In the short term, there is a need for immediately implementable solutions that leverage automation and artificial intelligence to make inspectors more efficient and effective decision makers. The validated method presented herein provides a way to rapidly localize and retrieve similar past inspection images using only a current inspection image, enabling bridge inspectors to make more informed decisions regarding the management of their structures. A composite similarity metric was introduced, which combines existing computer vision image similarity techniques with geolocation information. This method was applied to a data set of actual bridge deck inspection images to demonstrate that the CS scoring can be used successfully with or without accurate GPS information.

This study builds upon previous bridge image organization work to enable smart databases that are able to self-organize and provide historical inspection images of specific locations to engineers automatically (Zhang et al. 2022a). Full implementation will require further validation for the other bridge element categories in addition to the deck. Additional future work might include assigning GPS-deficient images with coordinates based on their relative similarity to other images in the data set. Further, this method would improve structural inspections for similar applications where

geospatial and visual imagery are preserved and utilized together in databases. For example, concrete causeway networks may be many kilometers long and experience failures that could be predicted based on visual imagery. A retrieval capability that fuses image similarity with geolocation may be beneficial for those and other similar large-scale structures. Researchers may also use this method as a part of other fully automated workflows, some of which were discussed in this paper.

### Data Availability Statement

Some or all data, models, or code that support the findings of this study are available from the corresponding author upon reasonable request. Some or all data, models, or code used during the study were provided by a third party. Direct requests for these materials may be made to the provider as indicated in the Acknowledgements.

### Acknowledgments

The authors would like to thank the Indiana Department of Transportation for providing the inspection images used in this study, and for providing permission to share these images and findings. The authors declare no funding or potential conflicts of interest with this work.

### References

- AASHTO. 2013. *AASHTO manual for bridge element inspection*. Washington, DC: AASHTO.
- Abdallah, A. M., R. A. Atadero, and M. E. Ozbek. 2022. "A state-of-the-art review of bridge inspection planning: Current situation and future needs." *J. Bridge Eng.* 27 (2): 03121001. [https://doi.org/10.1061/\(ASCE\)BE.1943-5592.0001812](https://doi.org/10.1061/(ASCE)BE.1943-5592.0001812).
- Alayoub, A., S. A. E. Rahim, S. Mustapha, D. Salam, A. Tehrani, and N. L. D. Khoa. 2022. "The application of machine learning to paint condition assessment using hyperspectral data." In *Proc., 12th Hyperspectral Imaging and Signal Processing: Evolution in Remote Sensing*, 1–6. New York: IEEE. <https://doi.org/10.1109/WHISPERS56178.2022.9955099>.
- Apple. 2021. "Recognizing people in photos through private on-device machine learning." Accessed February 15, 2023. <https://machinelearning.apple.com/research/recognizing-people-photos>.
- ASCE. 2021. "2021 report card for America's infrastructure: Bridges." Accessed February 3, 2023. <https://infrastructurereportcard.org/cat-item/bridges-infrastructure/>.
- Bianchi, E. L., N. Sakib, C. Woolsey, and M. Hebdon. 2022. "Bridge inspection component registration for damage evolution." *Struct. Health Monit.* 22 (Jun): 472–495. <https://doi.org/10.1177/14759217221083647>.
- Cao, Q., Y. Ying, and P. Li. 2013. "Similarity metric learning for face recognition." In *Proc., IEEE Int. Conf. on Computer Vision*, 2408–2415. New York: IEEE. <https://doi.org/10.1109/ICCV.2013.299>.
- Cha, Y. J., W. Choi, and O. Büyüköztürk. 2017. "Deep learning-based crack damage detection using convolutional neural networks." *Comput. Aided Civ. Inf.* 32 (5): 361–378. <https://doi.org/10.1111/mice.12263>.
- Choi, J., J. A. Park, S. J. Dyke, C. M. Yeum, A. Lenjani, and I. Bilonis. 2022. "Similarity learning to enable building searches in post-event image data." *Comput.-Aided Civ. Infrastruct. Eng.* 37 (2): 261–275. <https://doi.org/10.1111/mice.12698>.
- Chun, P. J., T. Yamane, and Y. Maemura. 2022. "A deep learning-based image captioning method to automatically generate comprehensive explanations of bridge damage." *Comput.-Aided Civ. Infrastruct. Eng.* 37 (11): 1387–1401. <https://doi.org/10.1111/mice.12793>.



- Deng, J., A. Singh, Y. Zhou, Y. Lu, and C. S. Lee. 2022. "Review on computer vision-based crack detection and quantification methodologies for civil structures." *Constr. Build. Mater.* 356 (Jun): 129238. <https://doi.org/10.1016/j.conbuildmat.2022.129238>.
- Dorafshan, S., and M. Maguire. 2018. "Bridge inspection: Human performance, unmanned aerial systems and automation." *J. Civ. Struct. Health Monit.* 8 (3): 443–476. <https://doi.org/10.1007/s13349-018-0285-4>.
- El-Naqa, I., Y. Yang, N. P. Galatsanos, R. M. Nishikawa, and M. N. Wernick. 2004. "A similarity learning approach to content-based image retrieval: Application to digital mammography." *IEEE Trans. on Med. Imaging* 23 (10): 1233–1244. <https://doi.org/10.1109/TMI.2004.834601>.
- FHWA (Federal Highway Administration). 2022. *Specifications for the national bridge inventory*. Washington, DC: FHWA.
- Gagliardi, V. F., G. Bella, R. Previti, and L. Menghini. 2022. "Automatic damage detection of bridge joints and road pavements by artificial neural networks ANNs." In *Proc., SPIE Remote Sensing*, 12268. Berlin: SPIE. <https://doi.org/10.1117/12.2636217>.
- Google. 2023. Google maps. Accessed January 12, 2023. <https://www.google.com/maps>.
- Graybeal, B. A., B. M. Phares, D. D. Rolander, M. Moore, and G. Washer. 2002. "Visual inspection of highway bridges." *J. Nondestruct. Eval.* 21 (3): 67–83. <https://doi.org/10.1023/A:1022508121821>.
- Hadsell, R., S. Chopra, and Y. LeCun. 2006. "Dimensionality reduction by learning an invariant mapping." In *Proc., Computer Society Conf. on Computer Vision and Pattern Recognition*, 1735–1742. New York: IEEE. <https://doi.org/10.1109/CVPR.2006.100>.
- Howard, A., et al. 2019. "Searching for MobileNetV3." In *Proc., IEEE/CVF Int. Conf. on Computer Vision*, 1314–1324. New York: IEEE. <https://doi.org/10.1109/ICCV.2019.00140>.
- Jiang, S., J. Zhang, W. Wang, and Y. Wang. 2023. "Automatic inspection of bridge bolts using unmanned aerial vision and adaptive scale unification-based deep learning." *Remote Sens.* 15 (2): 328. <https://doi.org/10.3390/rs15020328>.
- Jiang, X., A. K. Gramopadhye, and B. J. Melloy. 2004. "Theoretical issues in the design of visual inspection systems." *Theor. Issues Ergon. Sci.* 5 (3): 232–247. <https://doi.org/10.1080/1463922021000050005>.
- Kao, S. P., Y. C. Chang, and F. L. Wang. 2023. "Combining the YOLOv4 deep learning model with UAV imagery processing technology in the extraction and quantization of cracks in bridges." *Sensors* 23 (5): 2572. <https://doi.org/10.3390/s23052572>.
- Kim, I. H., H. Jeon, S. C. Baek, W. H. Hong, and H. J. Jung. 2018. "Application of crack identification techniques for an aging concrete bridge inspection using an unmanned aerial vehicle." *Sensors* 18 (6): 1881. <https://doi.org/10.3390/s18061881>.
- Kong, X., K. Li, Y. Zhang, and S. Das. 2022. "Bridge deck deterioration: Reasons and patterns." *Transp. Res. Rec.* 2676 (7): 570–584. <https://doi.org/10.1177/03611981221080140>.
- Krizhevsky, A., I. Sutskever, and G. E. Hinton. 2017. "Imagenet classification with deep convolutional neural networks." *Commun. ACM* 60 (6): 84–90. <https://doi.org/10.1145/3065386>.
- Ktena, S. I., S. Parisot, E. Ferrante, M. Rajchl, M. Lee, B. Glocker, and D. Rueckert. 2017. "Distance metric learning using graph convolutional networks: Application to functional brain networks." In Vol. 10433 of *Proc., Medical Image Computing and Computer Assisted Intervention—MCCAI 2017*, 469–477. New York: Springer. [https://doi.org/10.1007/978-3-319-66182-7\\_54](https://doi.org/10.1007/978-3-319-66182-7_54).
- Li, Z. J., K. Adamu, K. Yan, X. L. Xu, P. Shao, X. H. Li, and H. M. Bashir. 2022. "Detection of nut–bolt loss in steel bridges using deep learning techniques." *Sustainability* 14 (17): 10837. <https://doi.org/10.3390/su141710837>.
- Lim, R. S., H. M. La, and W. Sheng. 2014. "A robotic crack inspection and mapping system for bridge deck maintenance." *IEEE Trans. Autom.* 11 (2): 367–378. <https://doi.org/10.1109/TASE.2013.2294687>.
- Lin, J. J., A. Ibrahim, S. Sarwade, and M. Golparvar-Fard. 2021. "Bridge inspection with aerial robots: Automating the entire pipeline of visual data capture, 3D mapping, defect detection, analysis, and reporting." *J. Comput. Civ. Eng.* 35 (2): 04020064. [https://doi.org/10.1061/\(ASCE\)CP.1943-5487.0000954](https://doi.org/10.1061/(ASCE)CP.1943-5487.0000954).
- Lovelace, B., and L. Wells. 2018. *Improving the quality of bridge inspections using unmanned aircraft systems (UAS): Research Project Final Report*. St. Paul, MN: Minnesota DOT.
- Merry, K., and P. Bettinger. 2019. "Smartphone GPS accuracy study in an urban environment." *PLoS One* 14 (7): e0219890. <https://doi.org/10.1371/journal.pone.0219890>.
- Montes, K. M., J. Dang, J. Liu, and P. Chun. 2022. "Bridge status realization and management enhanced by UAV, SfM, and deep learning." In Vol. 2 of *Proc., European Workshop on Structural Health Monitoring*, 536–545. New York: Springer. [https://doi.org/10.1007/978-3-031-07258-1\\_55](https://doi.org/10.1007/978-3-031-07258-1_55).
- Nguyen, A., V. Gharehbaghi, N. T. Le, L. Sterling, U. I. Chaudhry, and S. Crawford. 2023. "ASR crack identification in bridges using deep learning and texture analysis." In Vol. 50 *Structures*, 494–507. Amsterdam, Netherlands: Elsevier. <https://doi.org/10.1016/j.istruc.2023.02.042>.
- Perry, B. J., Y. Guo, R. Atadero, and J. W. van de Lindt. 2020. "Streamlined bridge inspection system utilizing unmanned aerial vehicles (UAVs) and machine learning." *Measurement* 164 (Jun): 108048. <https://doi.org/10.1016/j.measurement.2020.108048>.
- Phares, B. M., G. A. Washer, D. D. Rolander, and B. A. Graybeal. 2004. "Routine highway bridge inspection condition documentation accuracy and reliability." *J. Bridge Eng.* 9 (4): 403–413. [https://doi.org/10.1061/\(ASCE\)1084-0702\(2004\)9:4\(403\)](https://doi.org/10.1061/(ASCE)1084-0702(2004)9:4(403)).
- Phillips, S., and S. Narasimhan. 2019. "Automating data collection for robotic bridge inspections." *J. Bridge Eng.* 24 (8): 04019075. [https://doi.org/10.1061/\(ASCE\)BE.1943-5592.0001442](https://doi.org/10.1061/(ASCE)BE.1943-5592.0001442).
- Schroff, F., D. Kalenichenko, and J. Philbin. 2015. "FaceNet: A unified embedding for face recognition and clustering." In *Proc., IEEE Conf. on Computer Vision and Pattern Recognition*, 815–823. New York: IEEE. <https://doi.org/10.1109/CVPR.2015.7298682>.
- Seo, J., E. Jeong, and J. P. Wacker. 2022. "Machine learning approach to visual bridge inspection with drones." In *Proc., Structures Congress 2022*. Reston, VA: ASCE. <https://doi.org/10.1061/9780784484180.013>.
- Shariff, A., J. F. Bonnefon, and I. Rahwan. 2017. "Psychological roadblocks to the adoption of self-driving vehicles." *Nat. Hum. Behav.* 1 (10): 694–696. <https://doi.org/10.1038/s41562-017-0202-6>.
- Spencer, B. F., V. Hoskere, and Y. Narazaki. 2019. "Advances in computer vision-based civil infrastructure inspection and monitoring." *Engineering* 5 (2): 199–222. <https://doi.org/10.1016/j.eng.2018.11.030>.
- Sutter, B., A. Lelevé, M. T. Pham, O. Gouin, N. Jupille, M. Kuhn, P. Lulé, P. Michaud, and P. Rémy. 2018. "A semi-autonomous mobile robot for bridge inspection." *Autom. Constr.* 91 (May): 111–119. <https://doi.org/10.1016/j.autcon.2018.02.013>.
- Wing, M. G., A. Eklund, and L. D. Kellogg. 2005. "Consumer-grade global positioning system (GPS) accuracy and reliability." *J. For.* 103 (4): 169–173. <https://doi.org/10.1093/jof/103.4.169>.
- Yamane, T., P. J. Chun, J. Dang, and R. Honda. 2023. "Recording of bridge damage areas by 3D integration of multiple images and reduction of the variability in detected results." *Comput.-Aided Civ. Infrastruct. Eng.* 38 (17): 2391–2407. <https://doi.org/10.1111/mice.12971>.
- Yang, L., R. Jin, L. Mummert, R. Sukthankar, A. Goode, B. Zheng, S. O. H. Hoi, and M. Satyanarayanan. 2010. "A boosting framework for visuality-preserving distance metric learning and its application to medical image retrieval." *IEEE Trans. Pattern Anal. Mach. Intell.* 32 (1): 30–44. <https://doi.org/10.1109/TPAMI.2008.273>.
- Yang, T., Y. Liu, Y. Huang, J. Liu, and S. Wang. 2023. "Symmetry-driven unsupervised abnormal object detection for railway inspection." *IEEE Trans. Ind. Inform.* 19 (12): 11487–11498. <https://doi.org/10.1109/TII.2023.3246995>.
- Yeum, C. M., S. J. Dyke, B. Benes, J. Hacker, R. Ramirez, A. Lund, and S. Pujol. 2019. "Post event reconnaissance image documentation using automated classification." *J. Perform. Constr. Facil.* 33 (1): 04018103. [https://doi.org/10.1061/\(ASCE\)CF.1943-5509.0001253](https://doi.org/10.1061/(ASCE)CF.1943-5509.0001253).
- Yu, W., and M. Nishio. 2022. "Multilevel structural components detection and segmentation toward computer vision-based bridge inspection." *Sensors* 22 (9): 3502. <https://doi.org/10.3390/s22093502>.
- Zhang, L., F. Yang, Y. D. Zhang, and Y. J. Zhu. 2016. "Road crack detection using deep convolutional neural network." In *Proc., IEEE Int.*

- Conf. on Image Processing (ICIP)*, 3708–3712. New York: IEEE. <https://doi.org/10.1109/ICIP.2016.7533052>.
- Zhang, X., et al. 2022a. “Machine-supported bridge inspection image documentation using artificial intelligence.” *Transp. Res. Rec.* 2677 (5): 720–736. <https://doi.org/10.1177/03611981221135803>.
- Zhang, X., C. Beck, A. Lenjani, L. Bonthron, A. Lund, X. Liu, S. J. Dyke, J. Ramirez, P. Baah, and J. Hunter. 2022b. “Enabling rapid large-scale seismic bridge vulnerability assessment through artificial intelligence.” *Transp. Res. Rec.* 2677 (2): 1354–1372. <https://doi.org/10.1177/03611981221112950>.