

**AUTOMATED BRIDGE INSPECTION IMAGE LOCALIZATION AND
RETRIEVAL BASED ON GPS-REFINED SIMILARITY LEARNING**

by

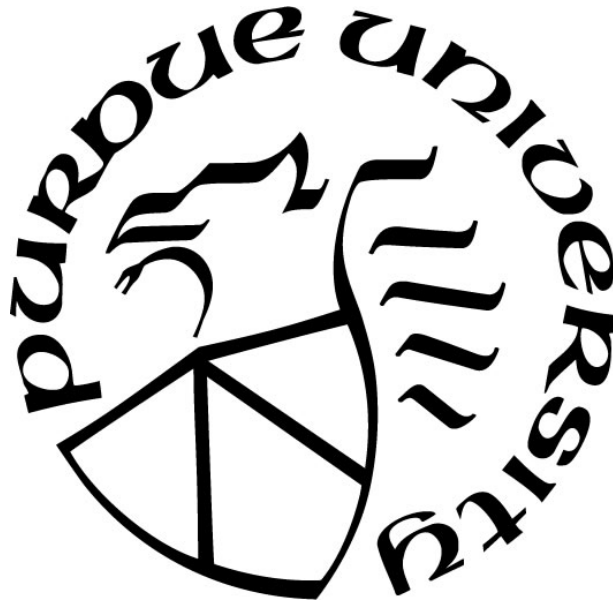
Benjamin E. Wogen

A Thesis

Submitted to the Faculty of Purdue University

In Partial Fulfillment of the Requirements for the degree of

Master of Science in Civil Engineering



Lyles School of Civil Engineering

West Lafayette, Indiana

May 2023

**THE PURDUE UNIVERSITY GRADUATE SCHOOL
STATEMENT OF COMMITTEE APPROVAL**

Dr. Shirley Dyke, Chair

Lyles School of Civil Engineering

Dr. Julio Ramirez

Lyles School of Civil Engineering

Dr. Randall Poston

Pivot Engineers

Approved by:

Dr. Dulcy Abraham

*Dedicated to my Parents, Jen and Steve, and my brother Connor.
And to my grandparents, Sandra, Joyce, and Eric.*

ACKNOWLEDGMENTS

I am incredibly grateful for the support and guidance of my advisor, Prof. Shirley Dyke, whom I had the pleasure of working with for the past three years as both an undergraduate and a graduate student. I would also like to thank Prof. Julio Ramirez and Dr. Randall Poston for kindly serving on my committee.

I am grateful for the mentoring of Dr. Jongseong Choi, whose papers introduced me to the topic of image similarity and who has met with me regularly to discuss the details of this research. I also acknowledge that much of the code used throughout this study was adapted from that which was provided by Dr. Choi.

Thank you to the rest of the Intelligent Infrastructure Systems Laboratory at Purdue University, past and present, for their continued feedback and support of my research. The members of the computer vision team, particularly Xin Zhang, Xiaoyu Liu, and Lissette Iturburu, served as valuable mentors as I refined my coding skills and learned about artificial intelligence. I am also appreciative of the JTRP predictive analytics team, particularly Manuel Salmeron and Nichole Criner, who allowed me to assist with their research and to tag along for many field visits and bridge inspections.

All of the bridge inspection photos utilized in this research were provided by the Indiana Department of Transportation, who has approved their use in this document. A derivation of this work has been submitted to a journal and is currently under review.

TABLE OF CONTENTS

LIST OF TABLES	7
LIST OF FIGURES	8
MATHEMATICAL NOTATION	9
ABSTRACT.....	10
1. INTRODUCTION.....	11
2. BACKGROUND.....	12
2.1 Automation in Bridge Inspection.....	12
2.2 Motivation for the Proposed Method.....	13
3. METHODOLOGY	15
3.1 Overview.....	15
3.2 Structural Element Classification	16
3.3 Image Similarity Training.....	17
3.4 GPS-Refined Composite Similarity.....	19
4. EXPERIMENTAL VALIDATION AND RESULTS.....	21
4.1 Initial Classification Step.....	21
4.2 Bridge Deck Dataset Overview	22
4.3 Labelling Definitions and Challenges.....	24
4.4 GPS Ground Truth Estimation.....	26
4.5 Similarity Training Setup.....	27
4.6 Composite Similarity Scoring and Ranking	28
5. DISCUSSION.....	32
5.1 Overall Performance	32
5.2 Composite Scoring Optimization.....	33
5.3 Practical Considerations.....	34
6. CONCLUSION.....	36
APPENDIX A. SUBSTRUCTURE CLASSIFIER RESULTS	37
APPENDIX B. DETAILED SIMILARITY LABELING	39
APPENDIX C. SAMPLE IMAGE QUERY RESULTS	41

REFERENCES	52
PUBLICATIONS.....	58

LIST OF TABLES

Table 1. Summary of the labelled dataset as divided into training and testing subsets.....	26
Table 2. Similarity ranking success for different proportions of database images having GPS...	31

LIST OF FIGURES

Figure 1. Workflow for the image localization and retrieval method.....	15
Figure 2. The classification schema applied in the developed method (adapted from Zhang et al. 2022).	16
Figure 3. Representation of the embedding process for (a) a similarly defined pair and (b) a dissimilarly defined pair.	18
Figure 4. The gathered dataset contains historical deck images for 40 bridges with varying GPS coordinate availability. Approximately half of the images (52%) have nonzero GPS.	22
Figure 5. Sample timeline of similar deck images taken from the dataset for a single bridge, showing EXIF metadata extracted from the image files and errors from the estimated ground truth.	23
Figure 6. Examples of labelling for typical bridge images.	25
Figure 7. Distribution of satellite-estimated ground truth distances for all image pairs in the validation dataset.	26
Figure 8. A sample of the loss history for 150 training epochs.	28
Figure 9. Similarity score distributions for all pairs in the testing dataset, shown for an example where $\beta = 0.30$, $dth = 20m$, and $pmin = -0.40$, for varying rates of GPS availability.....	30
Figure 10. Similarity rank distributions for all starting query images in the testing dataset, for the same parameters provided in Fig. 9.	30
Figure 11. Optimizing β for a database with moderate GPS error (0–50 meters).	34

MATHEMATICAL NOTATION

CS	composite similarity score for a given pair of images
D_W	Euclidean distance (L2-Norm) between two embedded image vectors
d	physical distance, in meters, between the coordinate metadata of a given pair of images
d_{th}	threshold distance, in meters
G_W	trained function used in the machine learning algorithm– in this publication, a Siamese convolutional neural network
GPS	latitude and longitude coordinates for a single image in the dataset
L	value of the contrastive loss function, to be minimized in the network training process
\mathcal{L}	set of inspection images belonging to a single bridge, for a single structural element of the bridge
m	margin used in the contrastive loss function
p_{min}	minimum penalty, representing the minimum score of the GPS component of the composite similarity score prior to being multiplied by the weight beta
SS	similarity score for a given pair of images without considering GPS information
W	weights of the parameters of the trained function, G_W
\vec{X}_i	vector representation of a single image in the dataset
\vec{x}_i	reduced dimension embedding vector for a single image in the dataset
Y	pairwise similarity training label
β	beta weight representing the proportion of the composite similarity score being contributed by the GPS component of the composite similarity equation

ABSTRACT

The inspection of highway bridge structures in the United States 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 as they are disorganized and unlabeled. Further, due to the lack of GPS metadata and visual ambiguity, it is often difficult for other inspectors to identify the location on the bridge where past images were taken. While many approaches are being considered toward fully- or semi-automated 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 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 dataset 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 to random image selection.

1. INTRODUCTION

The United States' inventory of highway bridge structures is a critical part of its national transportation system. According to the 2021 American Society of Civil Engineers' 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 United States 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- 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, I aim to develop a novel tool for bridge inspectors which enables the rapid recall of past bridge images to support their current decision-making. I develop a method which leverages deep machine learning and GPS coordinate information to compute a composite similarity (CS) metric between pairs of images which are then ranked and returned to the inspector. The approach is validated using a dataset 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.

2. BACKGROUND

2.1 Automation in Bridge Inspection

Current research aims to develop technologies using advanced computing and artificial intelligence to improve the efficiency and safety of bridge inspections. Towards this vision, researchers are developing systems to gather data autonomously or remotely using robotic agents (Phillips and Narasimhan 2019; Sutter et al. 2018; 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 Narasimhan, 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. 2023; Jiang et al. 2022), steel paint condition (Alayub et al. 2022), joint damage (Gagliardi et al. 2022), and for determining substructure vulnerability (Zhang et al. 2022).

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. While some techniques such as UAV-based inspections are already being utilized for special inspections, scaling these techniques to the entire NBI inventory 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. 2007).

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. 2023). While 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 (Zhang et al. 2022). 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 (Yamane et al. 2023).

2.2 Motivation for the Proposed Method

This thesis explores 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, I identify and leverage three extractable pieces of information, including: (1) the EXIF GPS coordinates; (2) the classification of the image in terms of its structural element category; and (3) the 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, while 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 pre-defined 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 (Spencer et al 2019; Yeum et al 2019; Zhang et al. 2022; 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 Siamese CNN (SCNN) and compared using available similarity metrics. Variations of this technique are used for applications in facial recognition (Cao et al. 2013; Shroff 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 post-reconnaissance analyses (Choi et al. 2021) and for detecting railroad track damage (Yang et al. 2023).

3. METHODOLOGY

3.1 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 the developed method must be able to consider both. The overall methodology is illustrated in Fig. 1.

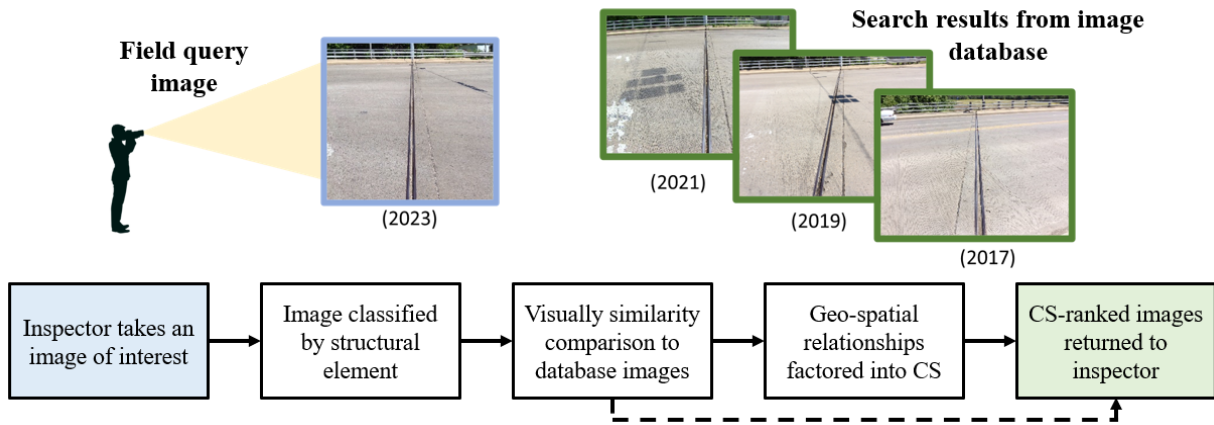


Figure 1. Workflow for the image localization and retrieval method

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, 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, while images with different learned features are mapped to spatially different vectors. The query image embedding is then compared to 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 is available for a given pair of images, the scores are adjusted using a novel composite similarity (CS) scoring metric which 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.

3.2 Structural Element Classification

Once the query image is captured or selected, it must be classified. Recent research conducted by our team 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 United States (Zhang et al. 2022; AASHTO 2013). Zhang et al. developed the classification schema shown in Fig. 2 and validated their method using a dataset of more than 11,000 images collected by inspectors, with an average accuracy of 94%.

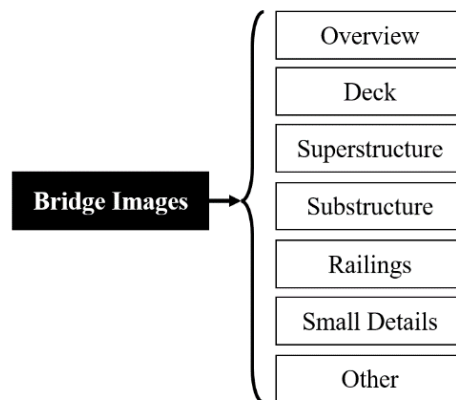


Figure 2. The classification schema applied in the developed method (adapted from Zhang et al. 2022).

This classifier is 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.

3.3 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, as the locations and angles of the deck images do not fit into any pre-defined 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):

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

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):

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

In this function, Y represents the labelled 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 labelled 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 pre-defined parameter that is always greater than zero. Hadsell et al. (2006) describe 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 Euclidian 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, as it is easy to visualize—however, any number of dimensions may be used for the embedding. In Fig. 3, the pair (a) of images 1 and 2 represents a similarly defined pair and is labelled with $Y = 0$, while the pair (b) of images 1 and 3 represents a dissimilarly defined pair and is labelled with $Y = 1$.

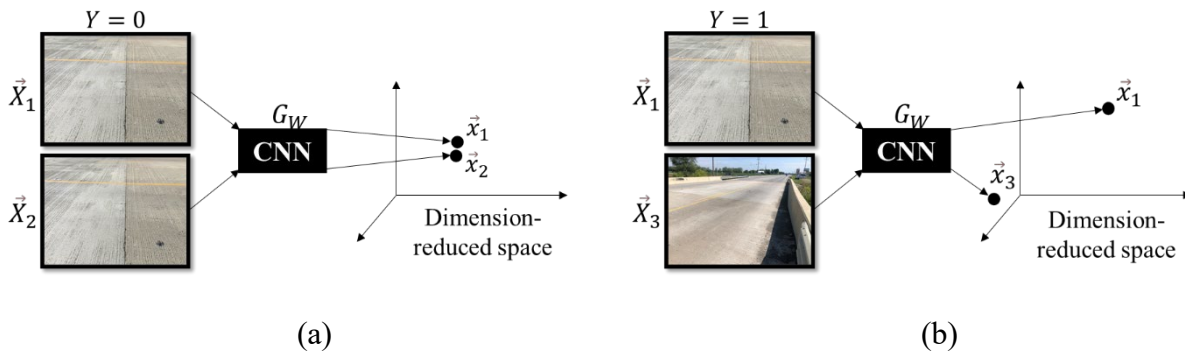


Figure 3. Representation of the embedding process for (a) a similarly defined pair and (b) a dissimilarly defined pair.

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

3.4 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). Since 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 SS defined as:

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

This function serves to normalize the Euclidian distance, D_W , to a value between 0 and 1. 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 1.

For this application, the SS concept is improved by considering the added benefit that image geolocation may have in the ranking process. A composite similarity (CS) score is thus developed which 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 as:

$$CS(\vec{X}_1, \vec{X}_2) = (1 - \beta) \left(\frac{1}{1 + D_W(\vec{X}_1, \vec{X}_2)} \right) + \beta * \max \left(p_{min}, \left(1 - \frac{d(GPS_1, GPS_2)}{d_{th}} \right) \right) \geq 0 \quad (4)$$

The first term is taken from Eq. 3 and reflects the output of the trained similarity model in a score ranging from 0 to 1. 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, in meters, between the coordinates GPS_1 and 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 dataset 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 1. 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 β 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 0 and 1. 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 dataset 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. Note that 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 β 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.

4. EXPERIMENTAL VALIDATION AND RESULTS

4.1 Initial Classification Step

As discussed in the methodology, image classification serves as an important pre-filtering step prior to evaluating image similarity. This step is completed in two levels by Zhang et al. (2022). In the first level, bridge images are successfully classified per Fig. 2 into: (1) overview images; (2) deck images; (3) superstructure images; (4) substructure images; (5) railing images; (6) very detailed images; and (4) other images. In addition to this first level of classification, several secondary-level classifiers are also developed and validated which further refine the overview, deck, superstructure, and substructure categories.

For my part in this task completed by Zhang et al. (2022), I am responsible for the secondary substructure classifier which further sorts the bridge substructure images into (1) bearing images; (2) pier overview images; (3) pier part images; (4) abutment overview images; and (5) abutment part images. To train and validate this classifier, the 2,841 total substructure images are labelled into the five sub-categories mentioned, and randomly split into training and validation sets at a 4:1 ratio for five consecutive trials. The base CNN chosen for this task is ResNet50 with an added SoftMax output layer. All images are resized to 512 X 512 X 3. Data augmentations are applied in the form of random rotations, shifts, zooms, and mirroring. Finally, a learning rate of 1×10^{-6} is used and the network is trained on a NVIDIA Titan X GPU (Zhang et al. 2022).

On average, the substructure classification accuracy achieved in validation is 86%, which is the poorest of all the classifiers developed. The first level of classification, for example, achieved an average accuracy of 94%. It is observed that the majority of the misclassifications in the substructure category occur between the part and overview categories (i.e., pier part images being classified as pier overview images, and vice versa). To illustrate this, an alternative secondary substructure schema was tested that combined the part and overview categories, thus considering only (1) bearing images; (2) pier images; and (3) abutment images. The network is then trained through the same procedure, and the average validation accuracy improved to 93%. We conclude

that there is a tradeoff to be considered in the classifier design between numerical accuracy and overall inspection utility, as to be determined by the inspectors and asset managers (Zhang et al. 2022). Several supporting figures and tables for this substructure classification task are provided in Appendix A.

4.2 Bridge Deck Dataset Overview

Given the quality and quantity of real-world inspection images used by Zhang et al. (2022), a subset of those images is chosen for validating the similarity portion of the method. Specifically, images from the deck category are 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). While forest canopies, mountains, and buildings may also impact the accuracy of recorded 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.

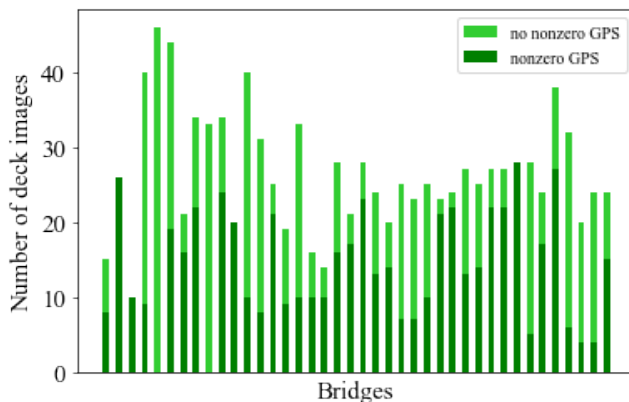


Figure 4. The gathered dataset contains historical deck images for 40 bridges with varying GPS coordinate availability. Approximately half of the images (52%) have nonzero GPS.

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 is available, as is the case for several of the bridges in the dataset. 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. 4 (a) and then be returned the images shown in Figs. 4 (b) through (f) corresponding to the previous five inspection cycles. These images have variations due to weather, camera type, deck deterioration, and maintenance actions, making the application of similarity a challenging task.

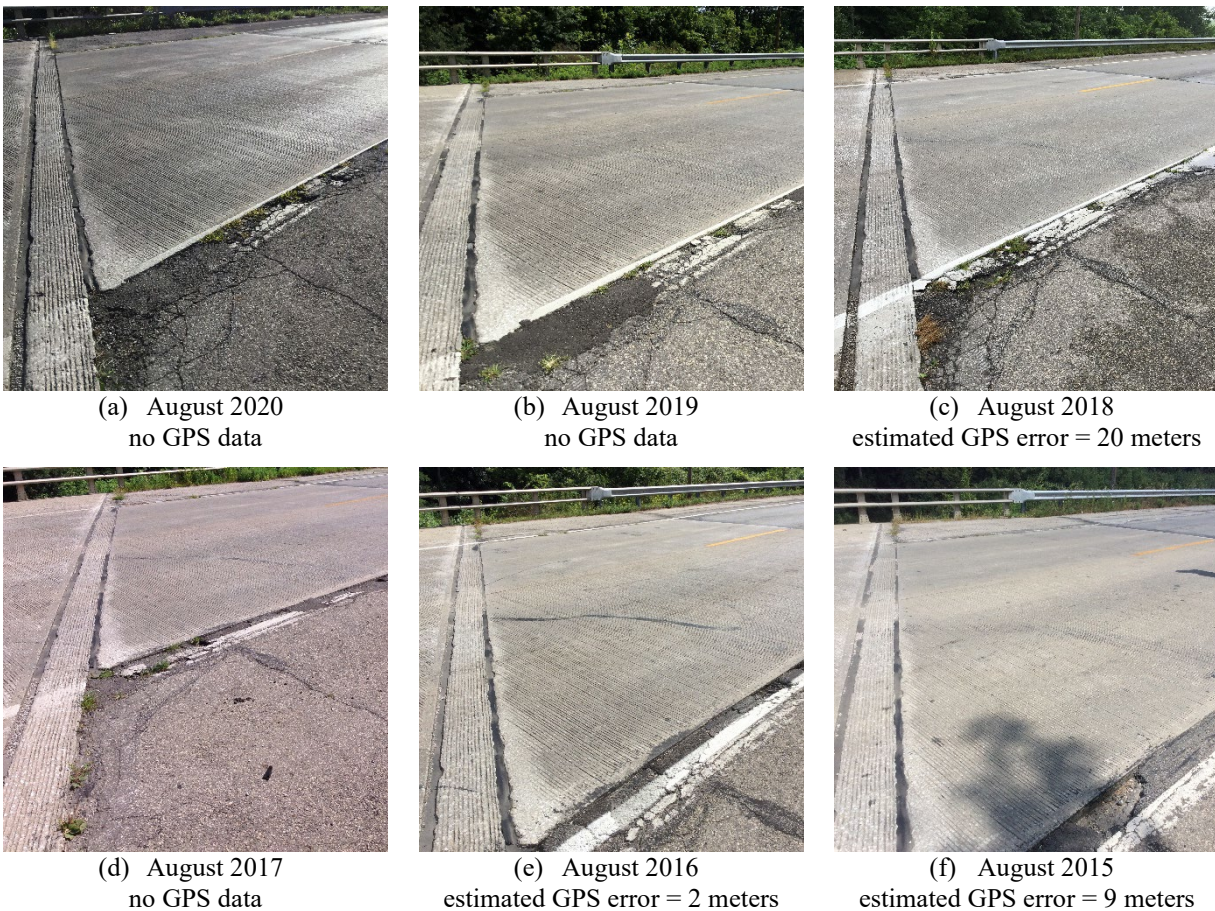


Figure 5. Sample timeline of similar deck images taken from the dataset for a single bridge, showing EXIF metadata extracted from the image files and errors from the estimated ground truth.

Fig. 5 is also representative of the quality of the GPS metadata in the dataset. GPS information may be present and reasonably accurate, present but inaccurate, or missing entirely. The image shown in Fig. 5 (e) is estimated to have accurate metadata, with the image EXIF coordinates being located only two meters from the estimated ground truth coordinates. Fig. 5 (c), however, has a significant estimated error of 20 meters. The images shown in Fig. 5 (a), (b), and (d) have no coordinate information. Since no ground truth exists for the image coordinates for the dataset used for training and validation, and since not all images have coordinates, the ground truth locations are estimated using Google satellite maps (Google, n.d.). 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.

4.3 Labelling Definitions and Challenges

While 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 labelling consistent across the entire dataset. 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 labelled 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 labelled with $Y = 1$. Image pairs taken from different bridge decks are not used in the training or evaluation of the model.

Several examples of challenging labelling situations are provided in Fig. 6, including (a) a deck joint; (b) a deck overview image; and (c) a close-up wearing surface image. Deck joints are some of the most frequently photographed components of a deck and contain many details. These images pose a challenge, as 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 labelled as true matches and share similar visual features. Likewise,

the wearing surface images like those shown in Fig. 6 (c) pose a challenge, as it is impossible to determine if they were taken at the same location or not.

The prescribed labelling definitions were strictly followed 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 labelled as similar pairs, even if it is unlikely that they are capturing the same exact location on the deck. This choice is made because these images serve the purpose of measuring the widest bridge cracks and are meant to be representative of the entire wearing surface of the structure at that point in time. Thus, these pairs are labelled as true matches so that inspectors can see the overall wearing surface condition in any given year as a result of their image search.

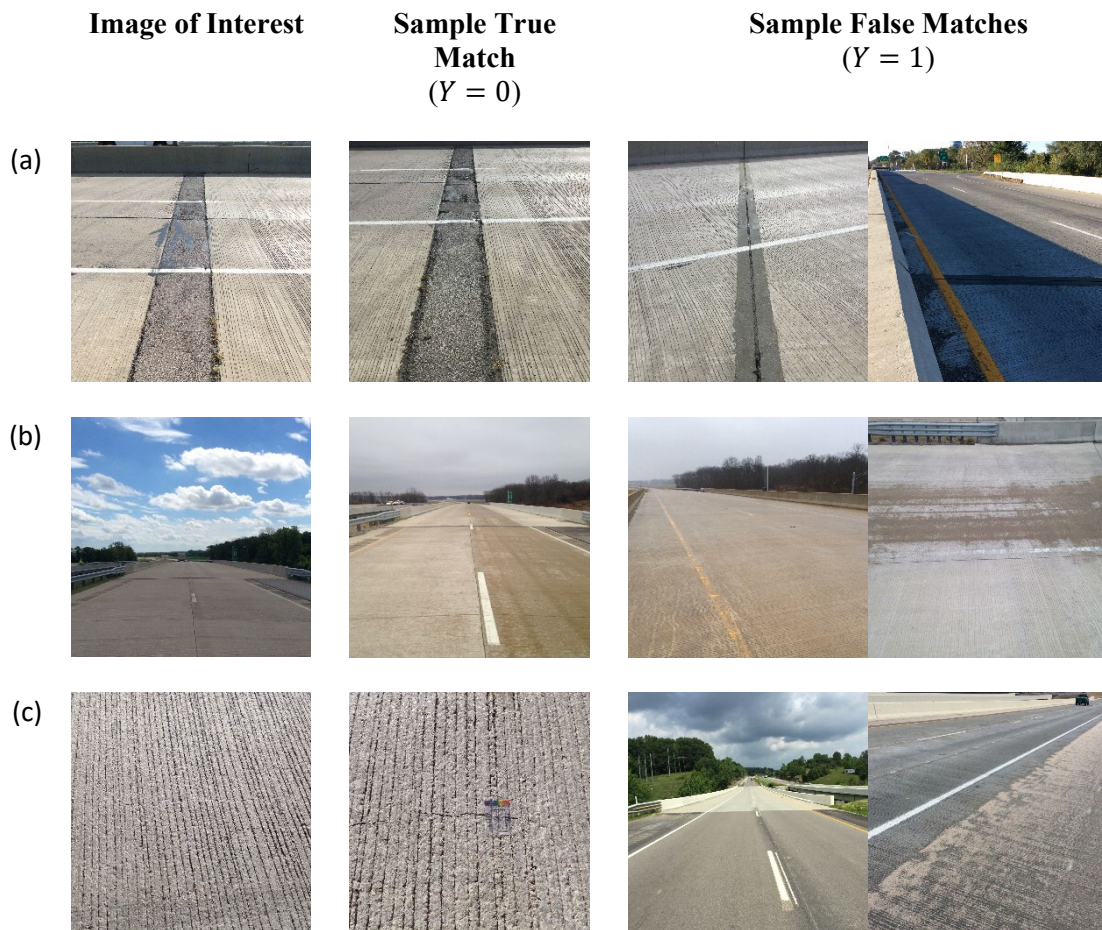


Figure 6. Examples of labelling for typical bridge images.

Overall, for the gathered 40-bridge dataset, 14,959 image pairings were manually labelled as prescribed in this section. The labelled dataset is summarized in Table 1, including the distribution of the training and testing sets described in the next section. A more detailed summary of the labels is included in Appendix B.

Table 1. Summary of the labelled dataset as divided into training and testing subsets.

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

4.4 GPS Ground Truth Estimation

As mentioned previously, to utilize the CS score developed in this paper, ground truth coordinates are estimated using satellite imagery from Google Maps (Google, n.d.). In this way, the point where the camera is located is estimated for each of the 195 testing images. While not perfect, I estimate the accuracy to be within 2 meters 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.

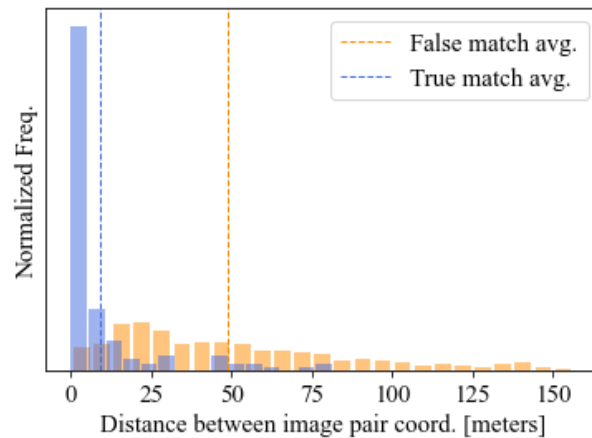


Figure 7. Distribution of satellite-estimated ground truth distances for all image pairs in the validation dataset.

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

4.5 Similarity Training Setup

The dataset is split randomly into a training dataset and a testing dataset at a 4:1 ratio, as summarized in Table 1. A total of 873 images belonging to 32 bridges are used to train the model, and 195 images belonging to 8 bridges are used in validation. While this may seem like a small dataset 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 are cropped to the maximum square size from the center point of the original image and are then downsized to 100 X 100 X 3. All three RGB-color channels are preserved, as the colors may contain valuable information for details such as lane markings. Several image sizes are tested, and downsizing the images was found to have little effect on the results.

In addition, data augmentation is applied in the form of random rotations of ± 10 degrees, contrast changes of $\pm 10\%$, brightness changes of $\pm 10\%$, and saturation changes of $\pm 10\%$ to avoid over-fitting (Choi et al. 2022). Zoom augmentations are not implemented such as to preserve the labelling intentions and to not accidentally omit similar deck features.

The network architecture chosen as the baseline model is MobileNetv3 (Howard et al. 2019). A batch size of 8 image pairs is used, and the learning rate is set to 5×10^{-4} . An NVIDIA TX TITAN X GPU is utilized to perform the training, and 150 training epochs are chosen empirically. The margin, m , is set to 2.0, and the output embedding size is chosen to be 10 based on trial and error. The resulting loss function given by Eq. 2 is calculated for both the training and validation set at each epoch, as shown in Fig. 8.

Overall, the training is successful, and the loss appears to converge to a minimum value after 150 epochs. The high noise and small loss improvement exhibited in the validation loss is due to the small size of the batches, which is necessary due to computing limitations. This behavior is acceptable, as 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 dataset. The training

requires approximately two hours on average for the given hardware and data processing specified in this section.

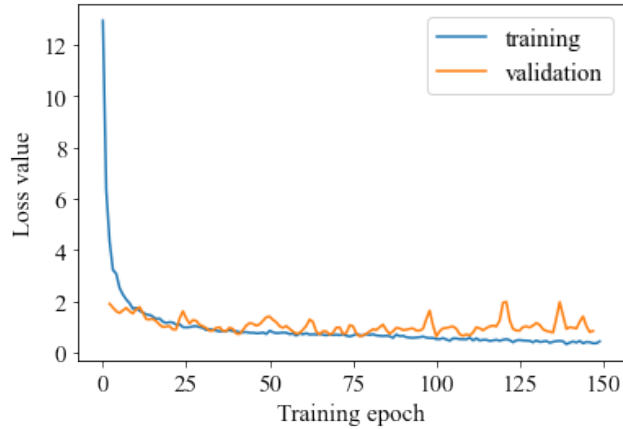


Figure 8. A sample of the loss history for 150 training epochs.

4.6 Composite Similarity Scoring and Ranking

Once the similarity model is trained, it is evaluated by computing CS scores (Eq. 4) for all of the 2,514 possible image pairings in the testing dataset and comparing the CS distributions for labelled true matches and labelled false matches. Further, for every query image, the position of the highest-ranked true match within a ranked list of same-bridge images is 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 is chosen for Eq. 4 as a reasonable value for this dataset, as most true match pairs have a distance falling below this value. Further, a maximum penalty of $p_{min} = -0.40$ and weight of $\beta = 0.30$ are chosen to illustrate the results. Using these parameters, the CS score is 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 (a) neglects all GPS; (b) samples GPS for 50% of images; (c) samples GPS for 75% of images; and (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) increases from 0.17 (a) to 0.36 (d) as the estimated ground truth GPS data is added to the dataset.

The specific images having coordinate information for each comparison are sampled at random. Recall that both images in a combination must have GPS information to have a nonzero

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

In a similar way, the performance is evaluated by ranking the CS results for each query image. Each image in the testing dataset is used once as the query image, and the CS score is 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. Since the inspector alone knows which precise feature they are interested in, they ultimately choose which image(s) 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. 10.

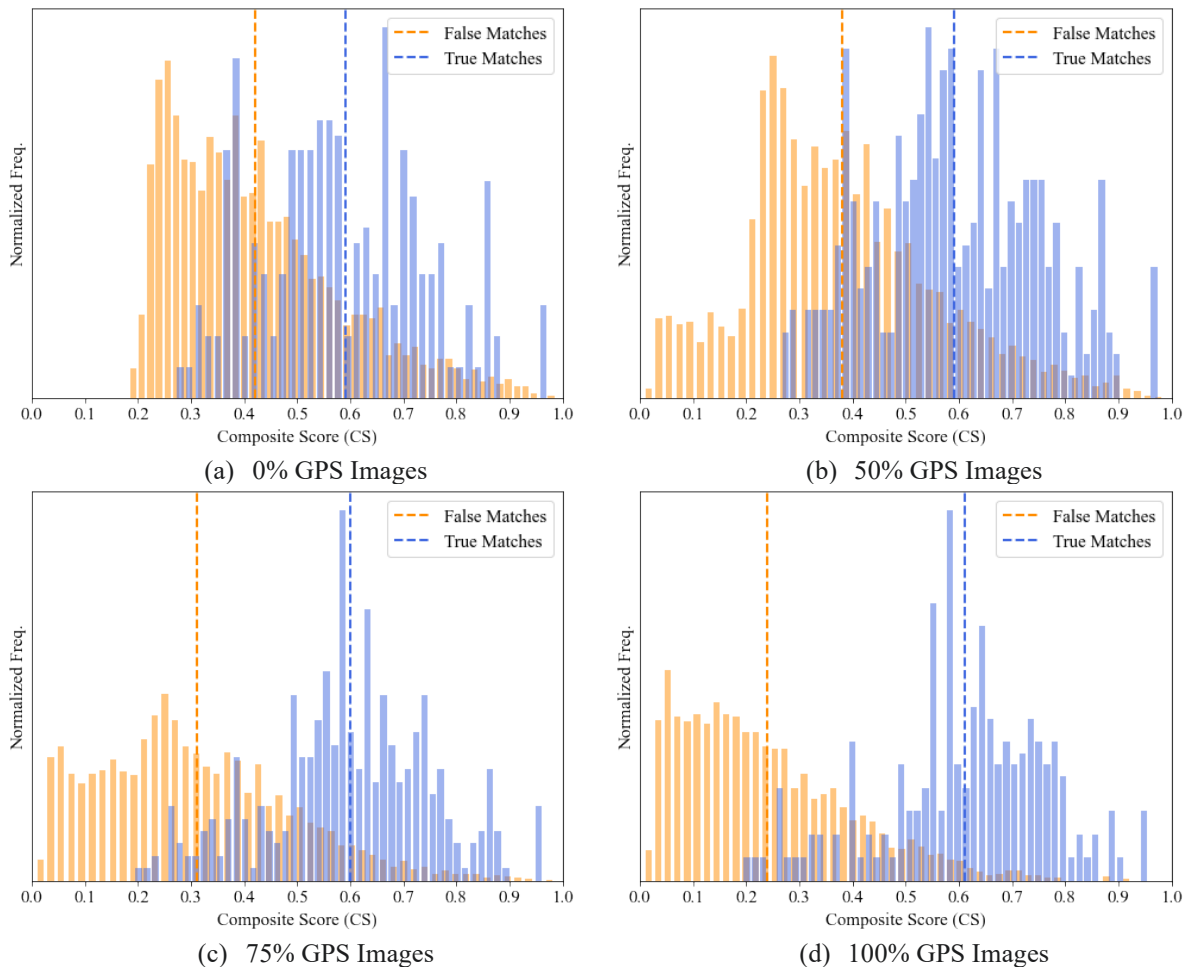


Figure 9. Similarity score distributions for all pairs in the testing dataset, shown for an example where $\beta = 0.30$, $d_{th} = 20m$, and $p_{min} = -0.40$, for varying rates of GPS availability.

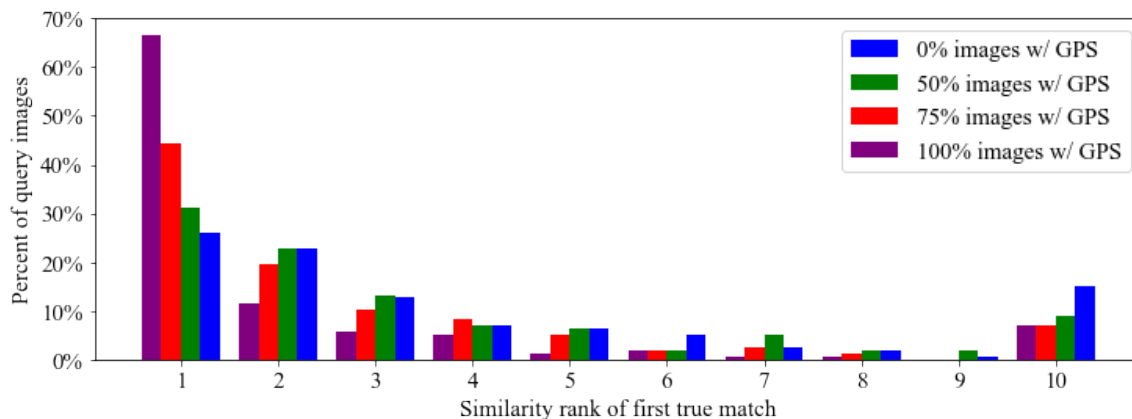


Figure 10. Similarity rank distributions for all starting query images in the testing dataset, for the same parameters provided in Fig. 9.

Fig. 10 illustrates that as more of the images in the database have quality GPS information, the true matches are able to be retrieved earlier in the CS-ranked list. While GPS does improve the results, it is not critical to the success of the method. In the case of 0% GPS images, for example, 61% of searches ranked a true match within the top three rankings. The performance of the similarity search is further summarized in Table 2 for several values of interest. The theoretical percentage of random selection is determined using the average number of images per bridge deck (~27), and the average number of true matches per image (~1). Images that have no true matches (36 of the 195 in the testing dataset) are not considered in this analysis. Table 2 shows that the proposed method performs favorably to random selection.

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

Images in dataset w/ GPS	Average True Match Similarity Rank	True match ranked 1 st	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%

5. DISCUSSION

5.1 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 are worse than previous implementations (Choi et al. 2022), and this result is likely due to the high variability in the true match definitions— this definition is decided to maximize utility for the inspectors, not to maximize performance indicators. Further, among true match pairs the background scenery varied significantly between inspection cycles due to seasonal changes and roadside 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 to 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 is available for the images in the database. From the results in Table 2, compared to 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 is considered and contains minimal error (as was assumed for this analysis) the results are improved, and an inspector has an eleven times greater likelihood of selecting a relevant image (true match) compared to random selection. Moreover, 90% of the time the inspector will find what they are looking for within the top 5 CS-ranked images. An example image query is provided in Appendix C, which shows the searched images and their ranked scores. In this example, the six true match images are ranked within the top eight positions out of the thirty-nine total comparisons.

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 dataset may contain more than 100 disorganized images.

5.2 Composite Scoring Optimization

The results assume that all GPS data is reasonably accurate, at least as much as the ground truth estimation process would allow. While 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 dataset 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 β selection for a dataset with artificial error is illustrated in Fig. 11. For this dataset, a random error between 0 and 50 meters is added to the distance between the ground truth coordinates for each image pair. Then, the method is 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 β values in Fig. 11 (a). Similarly, the percentages of all queries returning a true match within the top 3 rankings are provided in Fig. 11 (b). Due to the randomness in which images are assigned to GPS and which GPS comparisons are assigned with higher errors, this process is followed for 1000 iterations of β ranging from 0 to 1, to determine the value that, on average, maximizes the two chosen performance metrics.

The database manager can observe that when $\beta = 0$, the metrics converge to those of the 0% GPS dataset. This is because when no GPS data is available for one or both images in a pairing β is always taken to be zero. From the figure, the rank metric in Fig. 11 (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. 11 (b) is optimal for values of $\beta \approx 0.25$, regardless of GPS availability. Thus, for the hypothetical error-prone dataset 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.

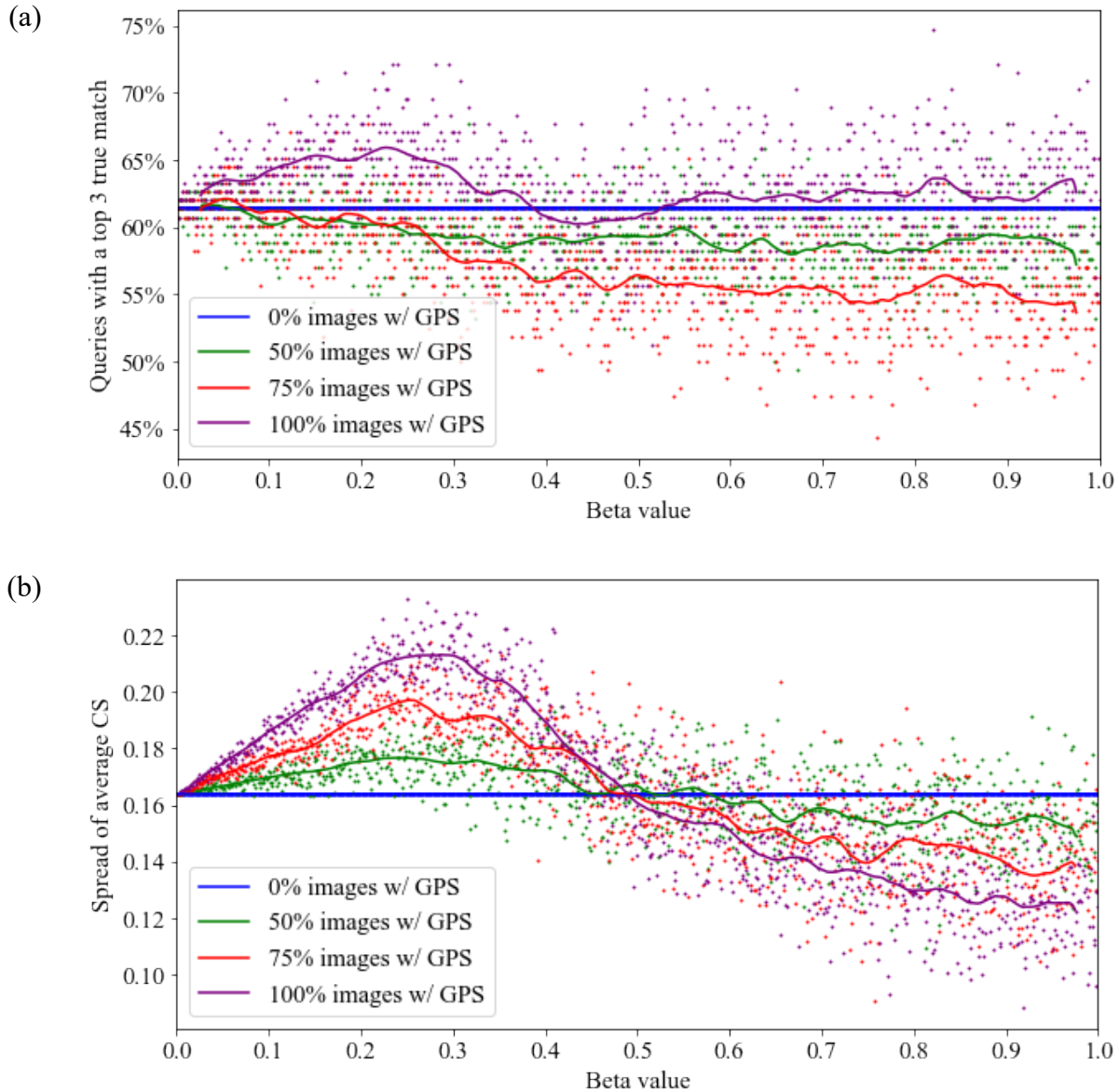


Figure 11. Optimizing β for a database with moderate GPS error (0–50 meters).

5.3 Practical Considerations

Based on the results, several recommendations are made 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 recommendations are: two alignment photos, one from each lengthwise side of

the bridge; two photos of each bridge joint and approach slab, one from each widthwise side of the bridge deck; one zoomed-in wearing surface photo representative of the entire deck; and 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 to the results presented herein, as this model is trained with highly non-uniform image locations and angles.

6. 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 which 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 (CS) metric is introduced which combines existing computer vision image similarity techniques with geolocation information. This method is applied to a dataset of actual bridge deck inspection images to demonstrate that the CS scoring can be used successfully with or without accurate GPS information.

Future work might include assigning GPS-deficient images with coordinates based on their relative similarity to other images in the dataset. 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 miles 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. For example, using this method, past images of the same bridge location and damage could be isolated, and deterioration rates could be estimated using automated damage detection and measurement algorithms.

APPENDIX A. SUBSTRUCTURE CLASSIFIER RESULTS

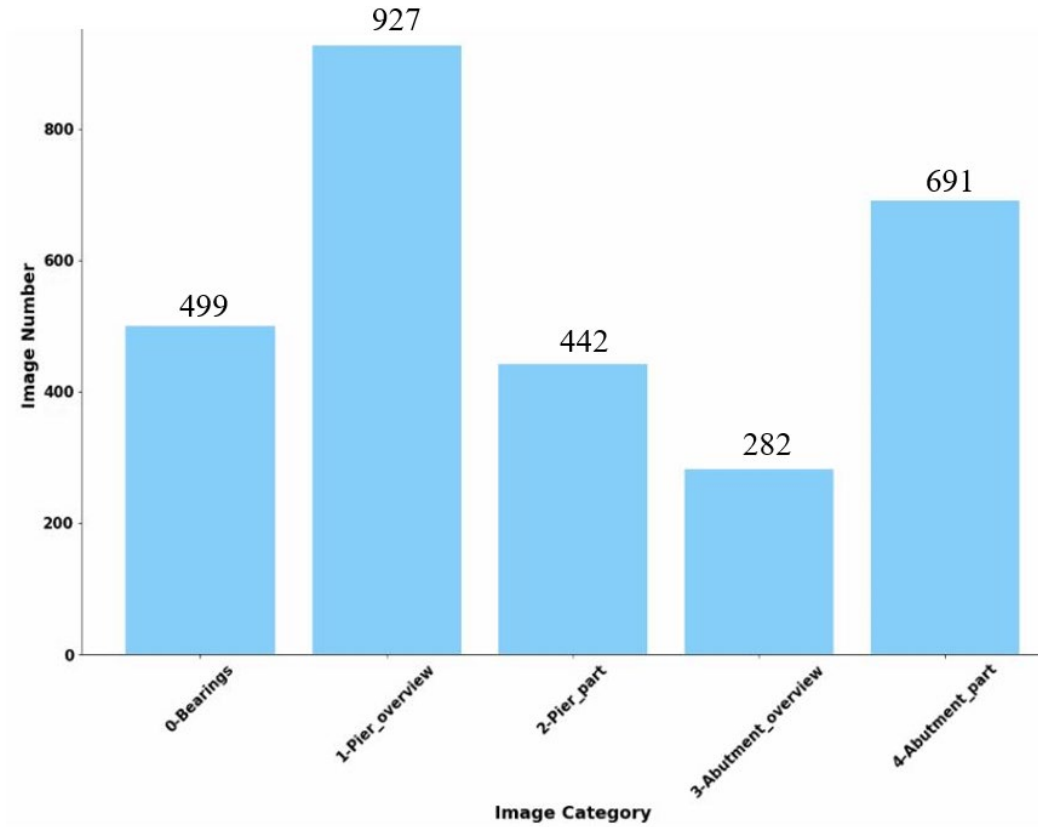


Figure A.1. Summary of the image ground truth labels for the secondary substructure image classifier for 2,841 total substructure images.

Table A.1. Validation results for each of the five random 4:1 dataset splits.

Dataset Split	Accuracy	Recall	Precision
1	0.844	0.819	0.826
2	0.877	0.865	0.863
3	0.865	0.861	0.852
4	0.851	0.827	0.826
5	0.842	0.821	0.820
Average	0.856	0.839	0.838

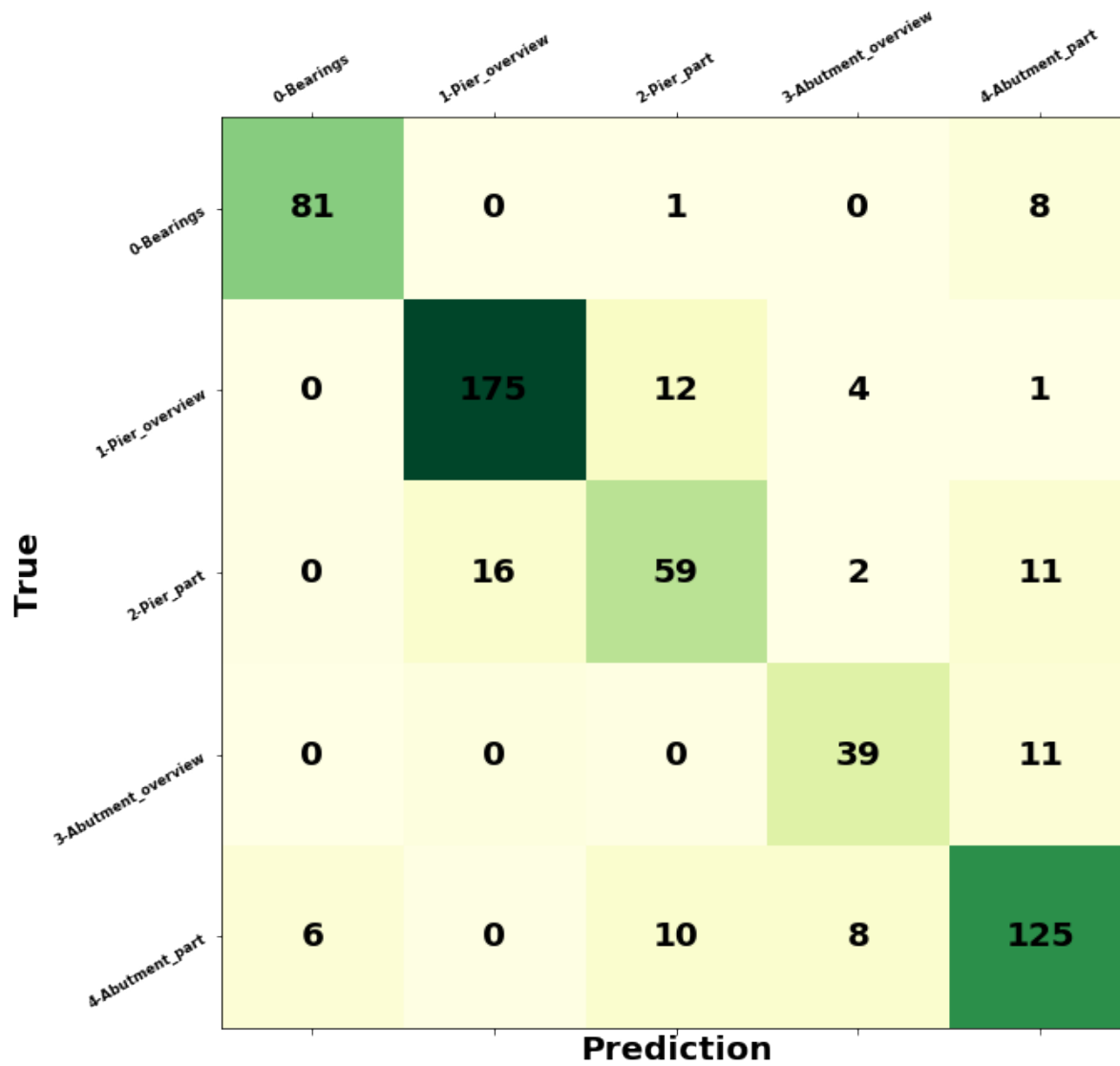


Figure A.2. Confusion matrix for one of the random dataset splits from Table A.1.

APPENDIX B. DETAILED SIMILARITY LABELING

Table B.1. Detailed summary of the training dataset labels.

Bridge No.	Dataset Split	No. Unique Images	No. Unique True Match Pairs	No. Unique False Match Pairs
1	Training	15	11	94
2	Training	26	48	277
3	Training	47	78	1,003
4	Training	44	75	871
5	Training	21	10	200
6	Training	34	34	527
7	Training	34	26	535
8	Training	20	21	169
9	Training	40	41	739
10	Training	31	26	439
11	Training	25	13	287
12	Training	19	17	154
13	Training	33	38	490
14	Training	16	23	97
15	Training	14	13	78
16	Training	28	22	356
17	Training	20	25	165
18	Training	28	48	330
19	Training	25	26	274
20	Training	23	16	237
21	Training	25	33	267
22	Training	23	19	234
23	Training	27	26	325
24	Training	27	29	322
25	Training	28	11	367
26	Training	28	41	337
27	Training	24	28	248
28	Training	38	38	665
29	Training	33	22	506
30	Training	20	4	186
31	Training	24	21	255
32	Training	33	19	509
Total		873	902	11,543

Table B.2. Detailed summary of the validation and testing dataset labels.

Bridge No.	Dataset Split	No. Unique Images	No. Unique True Match Pairs	No. Unique False Match Pairs
33	Validation/Test	10	5	40
34	Validation/Test	40	33	747
35	Validation/Test	24	17	259
36	Validation/Test	21	12	198
37	Validation/Test	24	21	255
38	Validation/Test	25	33	267
39	Validation/Test	27	27	324
40	Validation/Test	24	13	263
Total		195	161	2,353

APPENDIX C. SAMPLE IMAGE QUERY RESULTS



Figure C.1. Query image used to search the database, chosen as the test image with the most true match pairs (6). CS scoring parameters are set to be the same as in Figure 9 ($\beta = 0.30$, $d_{th} = 20\text{m}$, and $p_{min} = -0.40$).

Table C.1. Search results for the image shown in Figure C.1 (39 comparisons). The 6 true match images are all ranked within the top 8 CS-sorted values.



Image	Pairwise Label	Similarity Score	Distance Score	Composite Score (Rank)
	true match	0.78	0.96	0.83 (1)
	true match	0.72	0.96	0.79 (2)

Table C.1. continued.

Image	Label	Similarity Score	Distance Score	Composite Score (Rank)
	true match	0.59	0.85	0.66 (3)
	false match	0.66	0.65	0.66 (4)
	true match	0.56	0.65	0.59 (5)
	false match	0.62	0.49	0.58 (6)

Table C.1. continued.



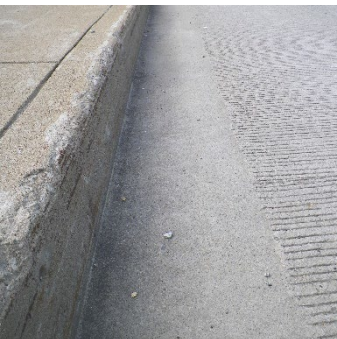

Image	Label	Similarity Score	Distance Score	Composite Score (Rank)
	true match	0.41	0.95	0.57 (7)
	true match	0.39	0.93	0.55 (8)
	false match	0.38	0.25	0.34 (9)
	false match	0.42	0.16	0.34 (10)

Table C.1. continued.

Image	Label	Similarity Score	Distance Score	Composite Score (Rank)
	false match	0.62	-0.40	0.32 (11)
	false match	0.58	-0.40	0.29 (12)
	false match	0.57	-0.40	0.28 (13)
	false match	0.56	-0.40	0.27 (14)

Table C.1. continued.





Image	Label	Similarity Score	Distance Score	Composite Score (Rank)
	false match	0.55	-0.40	0.27 (15)
	false match	0.33	0.13	0.27 (16)
	false match	0.55	-0.40	0.26 (17)
	false match	0.36	0.02	0.26 (18)

Table C.1. continued.





Image	Label	Similarity Score	Distance Score	Composite Score (Rank)
	false match	0.51	-0.40	0.24 (19)
	false match	0.43	-0.24	0.23 (20)
	false match	0.49	-0.40	0.22 (21)
	false match	0.49	-0.40	0.22 (22)

Table C.1. continued.





Image	Label	Similarity Score	Distance Score	Composite Score (Rank)
	false match	0.48	-0.40	0.22 (23)
	false match	0.47	-0.40	0.21 (24)
	false match	0.46	-0.40	0.20 (25)
	false match	0.40	-0.39	0.16 (26)

Table C.1. continued.




Image	Label	Similarity Score	Distance Score	Composite Score (Rank)
	false match	0.40	-0.40	0.16 (27)
	false match	0.40	-0.40	0.16 (28)
	false match	0.40	-0.40	0.16 (29)
	false match	0.40	-0.40	0.16 (30)

Table C.1. continued.



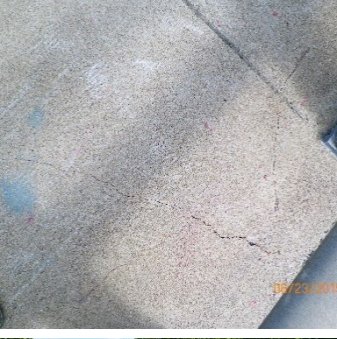
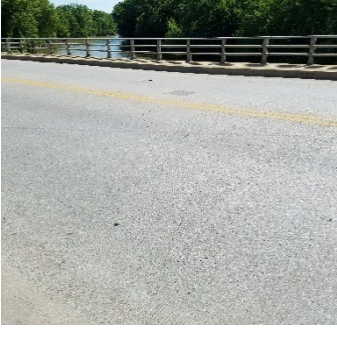
Image	Label	Similarity Score	Distance Score	Composite Score (Rank)
	false match	0.39	-0.40	0.15 (31)
	false match	0.37	-0.40	0.14 (32)
	false match	0.36	-0.40	0.13 (33)
	false match	0.34	-0.40	0.12 (34)

Table C.1. continued.






Image	Label	Similarity Score	Distance Score	Composite Score (Rank)
	false match	0.34	-0.40	0.12 (35)
	false match	0.34	-0.40	0.12 (36)
	false match	0.27	-0.40	0.07 (37)
	false match	0.26	-0.40	0.06 (38)

Table C.1. continued.

Image	Label	Similarity Score	Distance Score	Composite Score (Rank)
	false match	0.23	-0.40	0.04 (39)

REFERENCES

- AASHTO (American Association of State Highway and Transportation Officials). 2013. *AASHTO manual for bridge element inspection*. Washington, D. C.: AASHTO.
- Abdallah, A. M., Atadero, R. A., and Ozbek, M. E. 2022. "A state-of-the-art review of bridge inspection planning: current situation and future needs." *J. Bridge Eng.*, 27 (2).
[https://doi.org/10.1061/\(ASCE\)BE.1943-5592.0001812](https://doi.org/10.1061/(ASCE)BE.1943-5592.0001812).
- Alayoub, A., Rahim, S. A. E., Mustapha, S., Salam, D., Tehrani, A., and Khoa, N. L. D. 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. Rome, Italy: IEEE.
<https://doi.org/10.1109/WHISPERS56178.2022.9955099>.
- Apple. 2021. "Recognizing people in photos through private on-device machine learning." *Apple Machine Learning Research*. Accessed February 15, 2023.
<https://machinelearning.apple.com/research/recognizing-people-photos>.
- ASCE (American Society of Civil Engineers). 2021. "2021 report card for America's infrastructure: bridges" Accessed February 3, 2023.
<https://infrastructurereportcard.org/cat-item/bridges-infrastructure/>.
- Bianchi, E. L., Sakib, N., Woolsey, C., and Hebdon, M. 2022. "Bridge inspection component registration for damage evolution." *Struct. Health Monit.*, 22, 472–495.
<https://doi.org/10.1177/14759217221083647>.
- Cao Q., Ying, Y., and Li, P. 2013. "Similarity metric learning for face recognition." In *Proc. of the IEEE International Conference on Computer Vision*, 2408–2415. Sydney, Australia: IEEE. <https://doi.org/10.1109/ICCV.2013.299>.
- Choi, J., Park, J. A., Dyke, S. J., Yeum, C. M., Lenjani, A., and Bilonis, I. 2022. "Similarity learning to enable building searches in post-event image data." *Comput. Aided Civ. Inf.*, 37 (2), 261–275. <https://doi.org/10.1111/mice.12698>.

- Deng, J., Singh, A., Zhou, Y., Lu, Y., and Lee C. S. 2022. "Review on computer vision-based crack detection and quantification methodologies for civil structures." *Constr. Build. Mater.*, 356, 129238. <https://doi.org/10.1016/j.conbuildmat.2022.129238>.
- Dorafshan, S. and Maguire, M. 2018. "Bridge inspection: human performance, unmanned aerial systems and automation." *J. Civ. Struct. Health Monit.*, 8, 443–476. <https://doi.org/10.1007/s13349-018-0285-4>.
- El-Naqa, I., Yang, Y., Galatsanos, N. P., Nishikawa, R. M., and Wernick, M. N. 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, D. C.: FHWA.
- Gagliardi, V., Bella, F., Sansonetti, G., Previti, R., and Menghini, L. 2022. "Automatic damage detection of bridge joints and road pavements by artificial neural networks ANNs." In *Proc. SPIE Remote Sensing*, 12268. Berlin, Germany: SPIE. <https://doi.org/10.1117/12.2636217>.
- Google. (n.d.). *Google maps*. Accessed January 12, 2023. <https://www.google.com/maps>.
- Graybeal, B. A., Phares, B. M., Rolander, D. D., Moore, M., and Washer, G. 2002. "Visual inspection of highway bridges." *J. Nondestruct. Eval.*, 21, 67–83. <https://doi.org/10.1023/A:1022508121821>.
- Hadsell, R., Chopra, S., and LeCun, Y. 2006. "Dimensionality reduction by learning an invariant mapping." In *Proc. Computer Society Conference on Computer Vision and Pattern Recognition*, 1735–1742. New York, NY: IEEE. <https://doi.org/10.1109/CVPR.2006.100>.
- Howard, A., Sandler, M., Chu, G., Chen, L. C., Chen, B., Tan, M., Wang, W., Zhu, Y., Pang, R., Vasudevan, V., Le, Q. V. and Adam, H. 2019. In *Proc. of the IEEE/CVF International Conference on Computer Vision*, 1314–1324. Seoul, South Korea: IEEE. <https://doi.org/10.1109/ICCV.2019.00140>.

- Jiang, S., Zhang, J., Wang, W., and Wang, Y. 2023. "Automatic inspection of bridge bolts using unmanned aerial vision and adaptive scale unification-based deep learning." *Remote Sens.*, 2023, 15 (2), 328. <https://doi.org/10.3390/rs15020328>.
- Jiang, X., Gramopadhye, A. K., and Melloy, B. J. 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., Chang, Y. C., and Wang F. L. 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., Jeon, H., Baek, S. C., Hong, W. H., and Jung, H. J. 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., Li, K., Zhang, Y., and Das, S. 2022. "Bridge deck deterioration: reasons and patterns." *Transp. Res. Rec*, 2676 (7), 570–584. <https://doi.org/10.1177/03611981221080140>.
- Krizhevsky, A., Sutskever, I., and Hinton, G. E. 2017. "Imagenet classification with deep convolutional neural networks." *Commun. ACM*, 60 (6), 84–90. <https://doi.org/10.1145/3065386>
- Ktena, S. I., Parisot, S., Ferrante, E., Rajchl, M., Lee, M., Glocker, B., and Rueckert, D. 2017. "Distance metric learning using graph convolutional networks: application to functional brain networks." In *Medical Image Computing and Computer Assisted Intervention – MCCA 2017*, 10433, 469–477. Quebec City, Canada: Springer. https://doi.org/10.1007/978-3-319-66182-7_54.
- Li, Z. J., Adamu, K., Yan, K., Xu, X. L., Shao, P., Li, X. H., and Bashir, H. M. 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., La. H. M., and Sheng, W. 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., Ibrahim, A., Sarwade, S., and Golparvar-Fard, M. 2021. *J. Comput. Civ. Eng.*, 35 (2). [https://doi.org/10.1061/\(ASCE\)CP.1943-5487.0000954](https://doi.org/10.1061/(ASCE)CP.1943-5487.0000954).
- Lovelace, B. and Wells, L. 2018. *Improving the quality of bridge inspections using unmanned aircraft systems (UAS): Research Project Final Report*. St. Paul, MN: Minnesota Dept. of Transportation. <https://www.dot.state.mn.us/research/reports/2018/201826.pdf>.
- Merry, K. and Bettinger, P. 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., Dang, J., Liu, J., Chun, P. 2022. “Bridge status realization and management enhanced by UAV, SfM, and deep learning.” In *Proc. European Workshop on Structural Health Monitoring*, 2, 536–545. Palermo, Italy: Springer. https://doi.org/10.1007/978-3-031-07258-1_55.
- Nguyen, A., Gharehbaghi, V., Le, N. T., Sterling, L., Chaudhry, U. I., and Crawford, S. 2023. “ASR crack identification in bridges using deep learning and texture analysis.” *Struct.*, 50, 494–507, <https://doi.org/10.1016/j.istruc.2023.02.042>.
- Perry, B. J., Guo, Y., Atadero, R., and van de Lindt, J. W. 2020. “Streamlined bridge inspection system utilizing unmanned aerial vehicles (UAVs) and machine learning.” *Meas.*, 164, 108048. <https://doi.org/10.1016/j.measurement.2020.108048>.
- Phares, B. M., Washer, G. A., Rolander, D. D., and Graybeal, B. A. 2004. “Routine highway bridge inspection condition documentation accuracy and reliability.” *J. Bridge Eng.*, 9 (4). [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 Narasimhan, S. 2019. “Automating data collection for robotic bridge inspections.” *J. Bridge Eng.*, 24 (8). [https://doi.org/10.1061/\(ASCE\)BE.1943-5592.0001442](https://doi.org/10.1061/(ASCE)BE.1943-5592.0001442).
- Seo, J., Jeong, E., and Wacker, J. P. 2022. “Machine learning approach to visual bridge inspection with drones.” In *Proc. Structures Congress 2022*. Atlanta, Georgia: ASCE. <https://doi.org/10.1061/9780784484180.013>.
- Shariff, A., Bonnefon, J. F., and Rahwan, I. 2017. “Psychological roadblocks to the adoption of self-driving vehicles.” *Nat. Hum. Behav.*, 1, 694–696. <https://doi.org/10.1038/s41562-017-0202-6>.

- Schroff, F., Kalenichenko, D., and Philbin, J. 2015. "FaceNet: A unified embedding for face recognition and clustering." In *Proc. IEEE Conference on Computer Vision and Pattern Recognition*, 815–823. Boston, MA: IEEE. <https://doi.org/10.1109/CVPR.2015.7298682>.
- Spencer, B. F., Hoskere, V., and Narazaki, Y. 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., Lelevé, A., Pham, M. T., Gouin, O., Jupille, N., Kuhn, M., Lulé, P., Michaud, P., and Rémy, P. 2018. "A semi-autonomous mobile robot for bridge inspection." *Autom. Constr.*, 91, 111–119. <https://doi.org/10.1016/j.autcon.2018.02.013>.
- Wing, M. G., Eklund, A., and Kellogg, L. D. 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>.
- Yang, L., Jin, R., Mummert, L., Sukthankar, R., Goode, A., Zheng, B., Hoi S. O. H., and Satyanarayanan, M. 2010. "A boosting framework for visibility-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., Liu, Y., Huang, Y., Liu, J., and Wang, S. 2023. "Symmetry-driven unsupervised abnormal object detection for railway inspection." *IEEE Trans. Ind. Inform.* <https://doi.org/10.1109/TII.2023.3246995>.
- Yeum, C. M., Dyke, S. J., Benes, B., Hacker, J., Ramirez, R., Lund, A., and Pujol, S. 2019. "Post event reconnaissance image documentation using automated classification." *J. Perform. Constr. Facil.*, 33 (1). [https://doi.org/10.1061/\(ASCE\)CF.1943-5509.0001253](https://doi.org/10.1061/(ASCE)CF.1943-5509.0001253).
- Yu, W. and Nishio, M. 2022. "Multilevel structural components detection and segmentation toward computer vision-based bridge inspection." *Sensors*, 22 (9), 3502. <https://doi.org/10.3390/s22093502>.

Zhang, X., Wogen, B. E., Chu, Z., Dyke, S. J., Poston, R., Hacker, T., Ramirez, J., Liu, X., Iturburu, L., Baah, P., and Hunter, J. 2022. "Machine-supported bridge inspection image documentation using artificial intelligence." *Transp. Res. Rec.*

<https://doi.org/10.1177/03611981221135803>.

Zhang, X., Beck, C., Lenjani, A., Bonthron, L., Lund, A., Liu, X., Dyke, S. J., Ramirez, J., Baah, P., and Hunter, J. 2022. "Enabling rapid large-scale seismic bridge vulnerability assessment through artificial intelligence." *Transp. Res. Rec.*

<https://doi.org/10.1177/03611981221112950>.

PUBLICATIONS

- Criner, N. M., Salmeron, M., Zhang, X., Dyke, S. J., Ramirez, J. A., Wogen, B. E. 2023. “Predictive analytics for quantifying the long-term costs of defects during bridge construction.” (Joint Transportation Research Program Publication No. FHWA/IN/JTRP-2023/08). West Lafayette, IN: Purdue University. <https://doi.org/10.5703/1288284317615>
- Salmeron, M., Criner, N. M., Zhang, X., Dyke, S. J., Ramirez, J., Wogen, B., and Rearick, A. 2022. “Physics-based modelling of construction defects in concrete decks.” In Proc. *Eleventh International Conference on Bridge Maintenance, Safety and Management (IABMAS 2022)*. Barcelona, Spain: Taylor & Francis. <https://doi.org/10.1201/9781003322641>.
- Wogen, B. E., Choi, J., Zhang, X., Liu, X., Iturburu, L., and Dyke S. J. “Automated bridge inspection image localization and retrieval based on GPS-refined similarity learning.” *Under review*.
- Zhang, X., Wogen, B. E., Chu, Z., Dyke, S. J., Poston, R., Hacker, T., Ramirez, J., Liu, X., Iturburu, L., Baah, P., and Hunter, J. 2022. “Machine-supported bridge inspection image documentation using artificial intelligence.” *Trans. Res. Rec.* <https://doi.org/10.1177/03611981221135803>.
- Zhang, X., Wogen, B. E., Liu, X., Iturburu, L., Salmeron, M., Dyke, S. J., Poston, R., and Ramirez, J. A. 2023. "Machine-Aided Bridge Deck Crack Condition State Assessment Using Artificial Intelligence." *Sensors*, 23 (9), 4192. <https://doi.org/10.3390/s23094192>.