Combining Deep Learning and Robotics for Automated Concrete Delamination Assessment

E. McLaughlin[a], N. Charron[a], and S. Narasimhan[a]

[a]Department of Civil and Environmental Engineering, University of Waterloo, Canada
E-mail: em3mclau@uwaterloo.ca, ncharron@uwaterloo.ca, snarasim@uwaterloo.ca

Abstract –

Computer vision and robotics present tremendous opportunities for automating routine inspections of reinforced concrete bridges. One of the most critical aspects of these inspections is delamination assessment, as delaminations present immediate safety concerns due to falling concrete. Current methods of delamination assessment include hammer sounding and chain dragging, which are time consuming and difficult when accessibility is limited. Infrared technology presents an alternative method of assessing delaminations. In this work, a novel inspection method is proposed that uses an infrared camera combined with a convolutional neural network to automatically assess delaminations in infrared images. MobileNetV2 is implemented as an encoder with Deeplab V3 to perform pixel-wise segmentation in infrared images of delaminations.

The results show 74.5% mean intersection over union (mIoU) for predicting delaminated areas, which is comparable with the performance of this network architecture on benchmark data sets. Reviewing the predicted delamination areas also shows that the results accurately predict delamination locations, and accuracy limitations primarily exist in the fine outline details of the delamination. The automated delamination assessment method was also tested by mounting an upward facing thermal camera on a mobile ground robot to perform a bridge soffit inspection. The robotic scanning data set yielded a mIoU of 79.5% for delamination assessment. The increase in mIoU is likely due to the image data being better structured in the robotic images. This displays the ability to combine infrared imagery, convolutional neural networks, and unmanned mobile robots to meet case-specific accessibility needs for more accurate and time-efficient delamination assessment.

Keywords –

Convolutional neural network; Pixel-wise segmentation; automation; Delamination assessment; Bridge inspection

1 Introduction

Modern technologies in computer vision and robotics enable automation of routine tasks, with potential for improved efficiency in large-scale work. Routine concrete bridge inspection is one task where these technologies could provide great benefits [1]. In North America, there is demand for improved bridge inspections to more efficiently allocate the resources invested in bridge management. Routine inspections are performed by qualified inspectors following standardized guidelines (e.g., OSIM [2]). These inspections can be cumbersome and fraught with human error when access is limited due to long spans, water bodies, traffic conditions, and safety requirements. In addition, the results of routine inspections can be subjective and variable between inspectors [3].

One of the most important defects assessed during routine concrete bridge inspections is delamination. Concrete delamination can be described as a subsurface void in a concrete structure, which is caused by excessive internal stresses. The surface concrete next to the void is referred to as the delaminated concrete. Delaminations are critical, as they pose multiple safety hazards. The immediate hazard is that the mass of concrete below the void is likely to eventually fully detach and fall. This hazard is especially dangerous for bridges over high pedestrian or vehicular traffic areas.

In addition to being critical defects, delaminations are also one of the most difficult defects to assess during routine visual inspections because they are a subsurface defect that can seldom be detected by human vision. Currently, delaminations tend to be detected by manual process of hammer sounding or chain dragging to listen for hollow sounding concrete. These current sounding methods are time-consuming, inaccurate, and often not performed due to accessibility limitations. This makes delamination assessment a good candidate for an automated inspection method.

Infrared cameras combined with mobile robotic technologies present the opportunity for an efficient method of delamination assessment. The first step

towards developing this method is to use hand-held infrared cameras to assess delaminations. Delaminations can be identified in infrared images of concrete [4], as the delaminated concrete heats and cools at a different rate than the surrounding concrete. This occurs because the internal void acts as an insulating layer, making the delaminated concrete an independent, small mass [5]. This phenomenon makes delaminated areas often appear as a hot or a cold spot in an infrared image, depending on the time of day and environmental conditions. Detecting delaminations in infrared images is based on relative differences in thermal readings, so the absolute temperature of the delaminated areas is not critical for this study. Figure 1 shows examples of both hot and cold delaminations in infrared images of concrete bridge soffits.

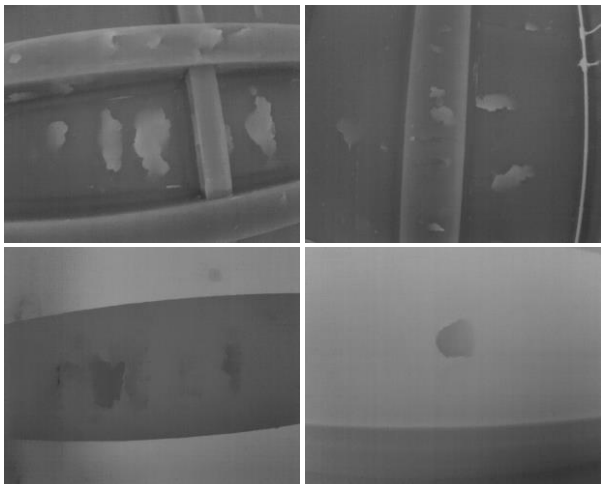

Figure 1. Examples of white-hot (top row) and black-cold (bottom row) delaminations in infrared images of concrete bridges.

Another advantage of using infrared cameras to detect delaminations is that this creates new possibilities for easy access to hard to reach bridge components. Many existing infrared cameras come in compact forms, ready to be mounted on an unmanned mobile robotic vehicle. This has the potential to eliminate many accessibility issues, as an infrared camera can be deployed on various vehicle types (aerial, ground, water) to account for the specific limitations of any bridge environment. For example, a bridge over water could have an upward facing infrared camera on a water vehicle to collect images. This proposed vision based infrared and robotic delamination assessment procedure provides a reasonable alternative in situations where traditional methods (e.g. hammer sounding) are not economically feasible due to special equipment and/or traffic control requirements.

Of all the advantages, perhaps the most useful advantage of using infrared imagery to assess delaminations is the ability to detect and localize the delaminations automatically using computer vision. In the computer vision community, there are two types of approaches that are used to automatically process image data for detecting objects of interest (delaminations being the objects of interest in this work): knowledge-driven methods and data-driven methods.

Knowledge-driven methods are based on the known characteristics of the object of interest. For example, this knowledge may include known colours and shapes of the objects of interest in the images. Knowledge based methods will often include human inputted rules for determining which features in images correspond to which objects. The second approach is data-driven methods, which rely on large quantities of labelled image data to train deep prediction networks. Once these networks have been trained to detect and localize the objects of interest based on the inputted data, they can then be used to detect and localize those objects on new data sets. These methods require large labelled data sets and significant processing resources to train the networks. Data-driven methods have become the method of choice for many computer vision based object recognition. These methods have consistently dominated computer vision competitions since a data-driven network first won the ImageNet challenge in 2012 [6].

For the assessment of concrete delaminations, knowledge-driven methods are difficult to implement successfully as delaminations have no definitive shape, and can show as a hot or cold area in the image depending on environmental conditions. This makes it difficult to determine a set of rules that will work across various delamination cases. Little research has been done to automatically assess delaminations using knowledge-driven methods, but some algorithms exist to automatically detect delaminations using a simple data-driven method (e.g., k-means clustering [7]). These existing methods are limited to specific bridge environments and often require some expert supervision.

Alternatively, convolutional neural networks (CNN) present an opportunity to use data-driven methods for fully automated detection. CNNs are considered state-of-the-art in computer vision as they currently dominate benchmarks such as [8], and can be leveraged to detect and localize objects of interest within images. Thus CNNs not only allow detection of images that show delaminations, but also provide the location of delaminated areas within each image. CNNs have yet to be applied to concrete bridge delamination assessment, but present a viable opportunity to improve upon existing results for delamination detection. This work proposes a CNN for semantic pixel-wise segmentation of delaminated areas within an infrared image of a reinforced concrete bridge.

The work presented herein can be divided into three main contributions. First, a versatile image labelling methodology is presented. This methodology fosters rapid pixel-wise labelling of concrete bridge defects in images to accelerate dataset building. Second, a novel approach is taken to concrete delamination detection and localization using infrared imagery combined with a convolutional neural network. Lastly, the described delamination detection methodology is demonstrated on a mobile ground robot. This final demonstration proves that the method presented herein can alleviate accessibility constraints in many bridge inspections. The methodology can be extended to aerial robots as well.

2 Related Works

Research work combining computer vision with concrete bridge inspection has primarily focused on automated crack detection and localization in visible spectrum images [9,10]. Some research has also been done on spall/delamination detection [11,12], but to a far lesser extent. With regards to delamination specifically, automated assessment via computer vision is still significantly limited. Since this work is focusing on delamination detection on concrete bridges using a CNN, this section will briefly present the work done to date on automated delamination detection as well as a the use of CNNs for general defect detection.

Automated detection of delaminations is fairly novel, as computer vision techniques applied to infrared imagery is required. Omar et al. [7] presented an automated procedure for detecting delaminations in concrete bridge decks using infrared thermography. They first created a thermal mosaic of the entire bridge deck from individual enhanced thermal images using a stitching algorithm. They then segmented the mosaic and identified objective thresholds using a k-means clustering method, to be able to identify delamination locations based on changes in heating patterns. This algorithm results in a condition map that effectively identifies delaminations within bridge decks.

A convolutional neural network (CNN) is a type of feed-forward neural network that utilizes the convolution operator in place of typical fully connected network layers. These CNNs are commonly applied to analyze inputs that have a grid-like data structure (e.g. pixel structure of images) [13]. Many practical computer vision applications require localization information within an image in addition to classification. Localization in this context means identifying which pixel areas in an image belong to a specific class (e.g., which pixels are part of a delamination). There are currently two CNN-based methods to achieve this: region based CNNs (R-CNN) and CNNs for semantic pixel-wise segmentation of images (semantic segmentation CNNs). R-CNNs improve on classification by adding a bounding box around the detected object to provide location information. Semantic segmentation CNNs provide structured grid-like data output, which provides pixel-level location accuracy within the image. Some researchers have begun to implement CNN algorithms for detection and/or localization of structural defects, and have shown promising results. However, much work is yet to be done for bridge inspections to incorporate the level of CNN technology that is currently available.

Region based CNNs (R-CNN) have been used more frequently than semantic segmentation CNNs. Cha et al. [14] developed a R-CNN algorithm for detecting cracks in high-resolution images. This method is computationally inefficient, as it uses a 256x256 pixel sliding window approach to classify each region as cracked/not cracked. The algorithm reported a high accuracy, but results were only tested on images of similar locations with a consistent background. This method was not proven to generalize to varying environments. Cha et al. [15] also used Faster R-CNN [16] for assessing five defect types (four types of steel defects and concrete cracking). This algorithm is efficient and showed promising test results for concrete cracks, but was limited to images of two bridges and a building complex. In addition, this study primarily focused on steel bridges rather than reinforced concrete bridges. Kim et al. [17] proposed a R-CNN as part of a complete bridge inspection solution. They used an unmanned aerial vehicle to collect images and combined it with a R-CNN for data processing and damage analysis. This work only studied visual images for crack assessment and did not investigate thermal defect analysis.

Zhang et al. [18] proposed a semantic segmentation CNN called CrackNet. CrackNet automates crack detection on 3D asphalt surfaces and maintains the shape of the original input image for pixel-wise classification/localization. This algorithm was successful for pixel-level detection of cracks but the results are only tested on asphalt. In addition, the algorithm is limited to crack labelling using 3D images from a high-quality PaveVision3D road inspection system. As a result, it is expected that this system cannot generalize to structural bridge inspections.

3 Methodology

The delamination assessment method proposed in this work uses a data-driven CNN for semantic pixel-wise segmentation. The first step to the proposed implementation is to collect a training data set large and diverse enough to provide high detection accuracy on a variety of concrete bridges. The following sections will

elaborate on the data collection and the labelling process. Finally, the implementation of the CNN will be presented in detail.

3.1 Image Data Set

A data-driven image segmentation approach is highly dependent on the size and quality of the image dataset provided. Therefore, collecting an adequate dataset to train, validate, and test a neural network is a critical step to the success of this method. There exist few public online data sets containing infrared images of delaminations, and no data sets where these images are labelled at a pixel-level to indicate delaminated areas. As a result, it is essential that a new infrared image data set be created and labelled appropriately.

The data set collected for this research consists of 500 infrared images taken of four reinforced concrete bridges in the cities of Kitchener and Waterloo, Ontario. The four selected bridges were in variable conditions. Images were taken at varying times of day, on different days, and under different weather conditions to ensure a variety in the data set. Of the 500 images, 261 contain some delamination, and 239 show no delamination in the concrete. All images were collected using a FLIR VUE Pro thermal camera. This camera can be mounted on small robots, including on unmanned aircraft systems, and can be remotely controlled. The camera has a 512x640 pixel resolution, and has a 7.5-13.5 micrometer spectral band. The images for this study were taken using the 'white hot' thermal camera setting (i.e., hot areas appear white in images, as shown in Figure 1).

The images that make up the dataset were collected in two ways: (1) manually and (2) automatically via unmanned ground robot. The manual method was used for the majority of the images in the dataset. In this case, the inspector held the camera. When the inspector suspected an area as delaminated, images were captured of that area using the FLIR camera phone application. The automatic image collection was done by mounting the FLIR camera in an upward facing configuration on a Husky unmanned ground robot (manufactured by Clearpath Robotics). Figure 2 shows this configuration including the robot and other sensors integrated into the system. This configuration allows the images to be captured at precise intervals and be synchronized with the other images taken on-board the robot. Furthermore, the exact position that the images were taken from can be determined based on the known coordinates of the robot.

The second collection method not only proves the additional benefits of integrating this defect detection method with other robotic sensors (e.g., GPS and visible spectrum cameras) but it also shows that this methodology can be used on any generic unmanned robot. As a result, this approach is suitable to address the accessibility problems with delamination assessment by utilizing the mobility of aerial, ground, and surface robotics.

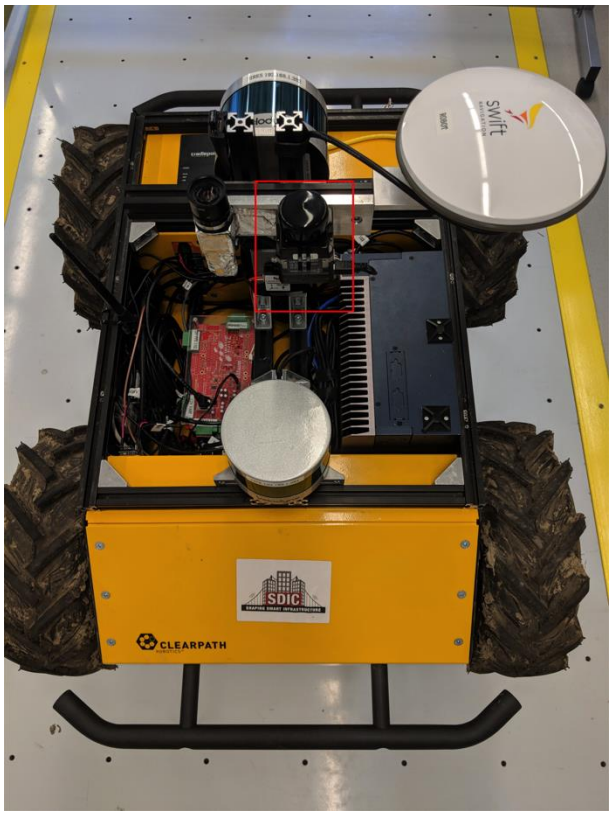

Figure 2. Husky unmanned ground vehicle with upward facing thermal camera.

3.2 Image Labelling

To train the proposed CNN to detect objects in images at a pixel level, each training image needs all the pixels to be labelled as one of the classes. In this case, only two classes are used: (1) sound concrete, and (2) delamination. Semantic pixel-wise image labelling is a tedious task, even with only two classes. To help with the labelling, a labelling graphical user interface (GUI) was created using MATLAB [19]. The goal of this GUI is to be generic enough that it can be used to label image pixels for any number of defect types on concrete bridges. This GUI makes use of basic image processing techniques to provide an initial estimate of the defect locations, and then provides tools to manually fine-tune the labelling.

Figure 3 shows the process used to provide the initial estimate of the pixel labels, prior to manual editing. First the image is converted to grayscale. This is required as grayscale thresholding is used to generate the initial delamination segmentation estimate. Even though most of the images used for this dataset were

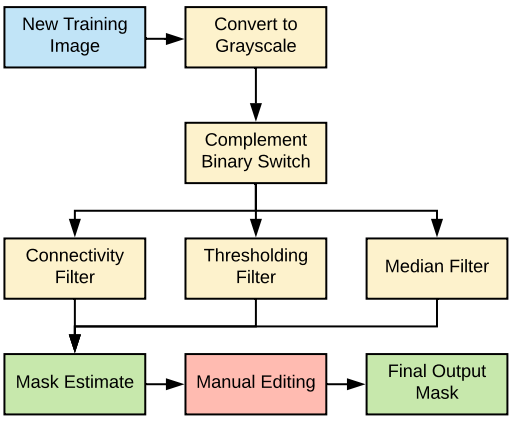

Figure 3. Flowchart of the image processing techniques used to simplify pixel-wise labelling.

collected in grayscale, this approach would generalize to any visible spectrum image or infrared images that use coloured heat gradients. The second step is to perform a complement binary switch (if required) to ensure the delamination consistently appears as back in the image mask. The three filters can then be applied including the thresholding, connectivity, and median filters. Each of these filters is applied using a sliding adjuster to allow easy editing of the three filters' sensitivities. Finally, a manual editing tool is used to remove unwanted areas or add missing areas.

Figure 4 shows the two main steps of the graphical interface for the image labeller. The top image shows the sliding filters. The bottom image shows the interface used for manually editing the initial mask.

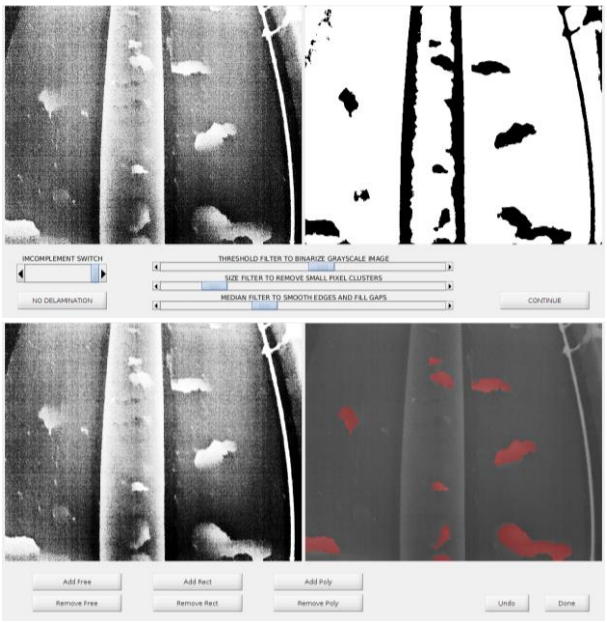

Figure 4. Image processing sliders to adjust delamination mask estimate (top) and GUI tools to manually adjust image mask (bottom).

These image processing techniques provide a good pixel-wise estimate of delaminated locations, but have several shortcomings. First, there tend to be other artifacts in the image that will be included during thresholding and cannot be removed by other filters. For example, girder edges tend to also heat/cool quicker than the surrounding concrete, giving them similar thermal patterns to a delaminated area. These edges are an example of an object requiring manual removal from the mask. Second, there can be some small gaps in the detected delamination area that cannot be fixed by the connectivity or median filters. This is an example of areas required to be manually added to the mask.

The described labelling approach allows for accurate, time-efficient labelling, which is necessary to be able to generate an adequate, well labelled data set. This approach has also been tested for labelling cracks, spalls, corrosion stains, and exposed reinforcement in concrete structures.

3.3 CNN Implementation

In order to properly assess delaminations in infrared imagery, both detection and pixel-level localization of the defect is required. In other words, it is not sufficient to say there is a delamination in the image, as the location of the delamination in the image is essential. The current state-of-the-art method for achieving this is a CNN algorithm. Many CNN algorithms exist to perform semantic pixel-wise segmentation, and often report exceptional results on standard segmentation data sets, such as Pascal VOC [8] or Cityscapes [20]. These datasets contain more classes, and more variable images compared to a dataset for delamination detection. Thus a similar quality of results should be expected when applied to delaminations, since the complexity of the data does not increase.

The network chosen for this application is MobileNetV2 [21] as a feature extractor with Deeplab V3 [22] to perform semantic pixel-wise segmentation. This network architecture was developed for mobile deployment, where time-efficiency is considered more important than accuracy. This is preferable as certain applications may benefit from real-time delamination assessment. For example, an inspector benefits more from real-time viewing of detected delamination locations on site than obtaining higher-resolution localization information about the exterior-boundary of the delaminated area.

The main building blocks of the network are from [21,22]. However, minor modifications were made to suit this specific application of this network to delamination assessment. The network input size is adjusted to be 512x640 to match the infrared image size. The last layer of the delamination prediction network is programmed to be a pixel-wise softmax activation layer,

which is given by the equation,

$$softmax(x)_i = \frac{\exp(x_i)}{\sum_{j=1}^{n} \exp(x_j)} \quad (1)$$

where x_i represents the network's calculated output score for a given class and n represents the number of classes [9]. In this case, $n = 2$ and the softmax activation is applied to predict each output pixel individually. This activation function restricts the output scores of each pixel to be between 0 and 1 so that the model can be trained using the categorical cross entropy (CCE) loss function, given by the equation,

$$J(\theta) = -\sum_{i=1}^{n} y_i \log(P(\theta)_i) \quad (2)$$

where θ represents the network weights, n is the number of possible classes, y_i is a binary classifier if the pixel does ($y_i = 1$) or does not ($y_i = 0$) belong to class i, and $P(\theta)_i$ is the network predicted likelihood that the pixel belongs to class i [9]. This combination of softmax output layer and CCE loss was used as it is generally considered best practice for multiple-class single-label classification. For this case it is also possible to use a sigmoid activation function with a binary cross entropy loss function, but this algorithm is intended to be scalable to include more defect classes in the future so the binary method was not preferred.

The network was trained using stochastic gradient descent [9], with a learning rate of 0.005, momentum of 0.9, and weight decay of 0.005/number of epochs. Image augmentation was used during training help prevent overfitting. Before each epoch, random transformations were applied to the training images only. These randomized image transformations include: flip horizontal, flip vertical, rotation, width and height shift, and zoom. This data augmentation ensured that the network would highly likely not see the exact same image twice during training.

4 Results

The collected 500 image data set was split into two main structures. The first structure consisted of randomly sorting the images into a 400 image training set, 50 image validation set, and 50 image test set. This yields a 80/10/10 dataset split. The second structure separated the images into a validation set consisting of images that were taken from the ground robot (Figure 2) under a bridge, and the training set consisted of all other images. This data split will be referred to as the robotic scanning split, where 62 images taken from a robot were used in the validation set and 436 of the remaining 438 images were used in the training set. Two images were left out to keep the number of training examples divisible by four (the desired batch size for this application).

The CNN algorithm for delamination assessment was reviewed for accuracy using the CCE loss (Equation 2) and mean intersection over union (mIoU) metrics. IoU is a common metric to compare segmentation results on benchmark datasets and is given by the equation,

$$IoU = \frac{TP}{TP + FP + FN} \quad (3)$$

where TP is true positives, FP is false positives, and FN is false negatives. In this application, positive is considered predicting a delamination and negative is considered predicting no delamination. CCE and mIoU were calculated and tracked at the end of each epoch for both the training and the validation set. Figure 5 shows the results of training on both dataset splits and Table 1 summarizes the final results.

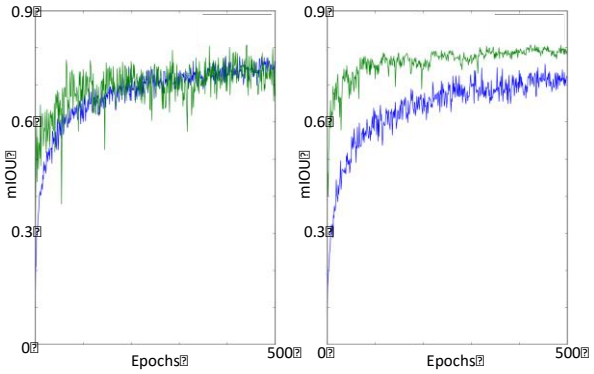

Figure 5. CNN results tracking mIoU during training for the 80/10/10 split (left) and robot scanning split (right). Blue represents training data and green represents validation data.

Table 1. Network Evaluation Metrics

Dataset	Split	CCE	mIoU
80/10/10	Train	0.0182	76.0%
	Val.	0.0163	72.7%
	Test	0.0169	74.5%
Robotic Scanning Split	Train	0.0166	71.1%
	Val.	0.0560	79.5%
Pascal VOC 2012	Benchmark	n/a	77.3% [8]

The results measured by CCE and mIoU show the network achieves comparable results for delamination assessment when compared to its performance on benchmark datasets. However, numbers cannot be compared to existing delamination assessment methods, as no fully automated methods exist on a similar infrared image dataset. To review the quality of the method, the output of the network was assessed on its

ability to predict delaminations for a typical inspection. Figure 6 shows delamination locations predicted by the network. The first three examples show results on images of bridge deck soffits that contain a various number of delaminated locations. The results show that the proposed methodology is capable of detecting nearly all delaminations, while avoiding false positives.

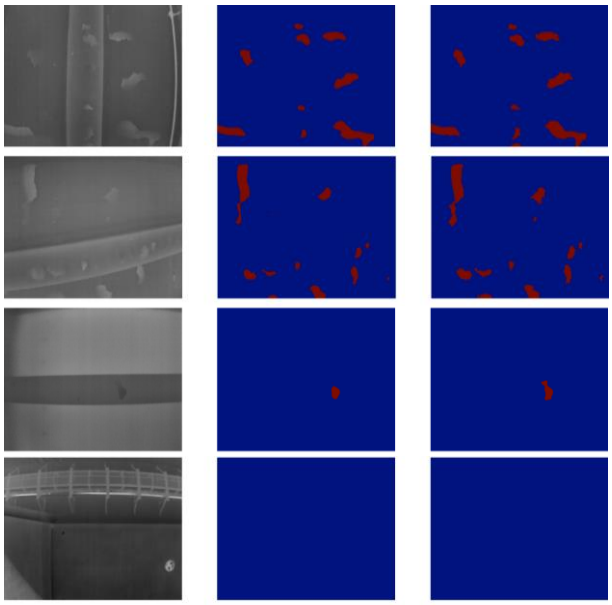

Figure 6. CNN example prediction results. Image (left), prediction (middle), true label (right).

The main source of error in these examples is the fine details. The edges of delaminations in the prediction are more rounded compared to the label, and small or thin delaminations are not well detected. This is a limitation based on the number of parameters used in the network and can be solved by adding more trainable parameters in the network. However, this limitation is not of great concern for concrete bridge inspection since precise knowledge of the delamination borders is not a requirement for bridge inspection. As discussed earlier, real-time visualization of the delamination assessment results is more valuable. In addition, traditional inspection procedures such as chain dragging and hammer sounding are not capable of precisely detecting edges and therefore this methodology is already a major improvement over current practice.

The fourth example in Figure 6 shows an example where there are a few potential areas that the network could falsely predict as a delamination. There is a scaffolding structure underneath the bridge, a small dark area and a small white area. In this case, the network successfully avoids false positives, showing that it has successfully learned some characteristics of a delamination and is not easily misled by similar artifacts.

It is also worth noting that the validation set of the robotic scanning split achieved the highest mIoU rating. This is most likely because the images taken by the robotic scan contained a large number of fairly clear delaminations. This is evidence that the quality of the image can influence the ability of the network to predict delamination locations.

5 Conclusion and Future Work

This work presents a novel automated concrete delamination assessment methodology using infrared imagery, robotics, and CNNs. A versatile pixel-wise defect labeler was first developed specifically for concrete bridge defects and was shown to help expedite the pixel-wise labelled image dataset creation. MobileNetV2 (Deeplab V3 implementation) was used to perform pixel-wise segmentation on infrared images of reinforced concrete bridges. This method was able to achieve high performance when evaluated on a data set of 500 infrared images by accurately detecting and locating delaminations under varying bridge environments. Quantitatively, the mIoU score for delamination assessment was shown to be comparable to what this network achieved on the Pascal VOC 2012 benchmark dataset. Qualitatively, the network was shown to predict delaminated areas correctly and avoid areas that have high potential to be classified as a false positive. The majority of the errors lowering the mIoU score are due to fine details around the edges of the delaminations which is of little concern to bridge inspections.

The network was also shown to be able to combine with a mobile ground robot (or any other unmanned mobile robotic platform). 62 images of a concrete bridge soffit and girder were taken by an upward facing compact infrared camera mounted on the robot and delaminated areas were predicted by the network. This test achieved high mIoU, which illustrates the ability of the delamination assessment network to integrate with robotic inspection. This is highly beneficial, as different unmanned robots can be used to accommodate specific accessibility limitations. Overall, this methodology presents significant improvements over state-of-the-art bridge inspection.

Future work will involve collecting a larger dataset to improve results, but a more complex network is likely not necessary as finer details are redundant for delamination assessment. In addition to more data, a quantitative comparison will be performed between the automated robotic assessment results and results from traditional sounding methods. Lastly, this methodology will be extended to other concrete bridge defects such as spalls, cracks, corrosion stains, and exposed reinforcing bar in visible spectrum images.

References

[1] Koch C., Georgieva K., Kasireddy V., Akinci B and Fieguth P. A review on computer vision based defect detection and condition assessment of concrete and asphalt civil infrastructure. *Advanced Engineering Informatics*, 29(2):196–210, 2015.

[2] Ministry of Transportation Ontario, St. Catharine's, ON. *Ontario Structural Inspection Manual (OSIM)*, 2008.

[3] Graybeal B.A., Phares B.M., Rolander D.D., Moore M. and Washer G. Visual inspection of highway bridges. *Journal of Nondestructive Evaluation*, 21(3):67–83, 2002.

[4] Yang X., Li H., Yu Y., Luo X., Huang T. and Yang X. Automatic Pixel-Level Crack Detection and Measurement Using Fully Convolutional Network. *Computer-Aided Civil and Infrastructure Engineering*, 33(12):1090–1109, 2018.

[5] Clemena G.G. and McKeel W.T. Detection of Delamination in Bridge Decks with Infrared Thermography. *Transportation Research Recorc No. 664, Bridge Engineering, Volume 1*, pages 180–182, 1978.

[6] Krizhevsky A., Sulskever I. and Hinton G.E. ImageNet Classification with Deep Convolutional Neural Networks. *Advances in Neural Information and Processing Systems (NIPS)*, 60(6):84–90, 2012.

[7] Omar T., Nehdi M.L. and Zayed T. Infrared thermography model for automated detection of delamination in RC bridge decks. *Construction and Building Materials*, 168(April):313–327, 2018.

[8] Everingham M., Van Gool L., Williams C.K.I., Winn J. and Zisserman A. The PASCAL Visual Object Classes (VOC) Challenge 2012. *International Journal of Computer Vision*, 88(2):303-338, 2010.

[9] Torok M.M., Golparvar-Fard M. and Kochersberger K.B. Image-Based Automated 3D Crack Detection for Post-disaster Building Assessment. *Journal of Computing in Civil Engineering*, 28(5):A4014004-1, 2014.

[10] Jahanshahi M.R., Masri S.F., Padgett C.W. and Sukhatme G.S. An innovative methodology for detection and quantification of cracks through incorporation of depth perception. *Machine Vision and Applications*, 24(2):227-241, 2013.

[11] German S., Brilakis I. and Desroches R. Rapid entropy-based detection and properties measurement of concrete spaling with machine vision for post-earthquake safety assessments. *Advanced Engineering Informatics*, 26(4):846-858, 2012.

[12] Adhikari R.S., Moselhi O. and Bagchi A. A study of image-based element condition index for bridge inspection. In *Proceedings of ISARC 2013 – 30th International Symposium on Automation and Robotics in Construction and Mining, Held in Conjunction with the 23rd World Mining Congress*, pages 345-356, Montreal, Canada, 2013.

[13] Goodfellow I., Benigo Y. and Courville A. *Deep Learning*. MIT Press, 2016.

[14] Cha Y.J., Choi W. and Büyüköztürk O. Deep Learning-Based Crack Damage Detecion Using Convolutional Neural Networks. *Computer-Aided Civil and Infrastructure Engineering*, 32(5):361–378, 2017.

[15] Cha Y.J., Choi W., Suh G., Mahmoudkhani S. and Büyüköztürk O. Autonomous Structural Visual Inspection Using Region-Based Deep Learning for Detection Multiple Damage Types. *Computer-Aided Civil and Infrastructure Engineering*, 33(9):731–747, 2018.

[16] Ren S., He K., Girshick R. and Sun J. Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks. In *Advances in Neural Information Processing Systems 28*, pages 91–99, Montreal, Canada, 2015.

[17] Kim I.H., Jeon H., Baek S.C., Hong W. and Jung H.J. Application of crack identification techniques for an aging concrete bridge inspection using unmanned aerial vehicle. In *Sensors*, 18(6):1881, Basel, Switzerland, 2018.

[18] Zhang A., Wang K.C.P., Li B., Yang E., Dai X., Peng Y., Fei Y., Liu Y., Li J.Q. and Chen C. Automated pixel-level pavement crack detection on 3d asphalt surfaces using a deep-learning network. *Computer-Aided Civil and Infrastructure Engineering*, 32(10):805–819, 2017.

[19] MATLAB. Version 9.2.0 (R2017a). The MathWorks Inc., Natick, Massachusetts, 2017.

[20] Cordts M., Omran M., Ramos S., Rehfeld T., Enzweiler M., Benenson R., Franke U., Roth S. and Schiele B. The cityscapes dataset for semantic urban scene understanding. In *Proceedings of the IEEE Conference on Computer Vision and Patter Recognition (CVPR), Las Vegas, USA*, 2016.

[21] Sandler M., Howard A.G., Zhu M., Zhmoginov A. and Chen L.C. Mobile networks for classification, detection and segmentation. *CoRR*, abs/1801.04381, 2018.

[22] Chen L.C., Papandreou G., Schroff F. and Adam H. Rethinking atrous convolution for semantic image segmentation. *CoRR*, abs/1706.05587, 2017.