

ORIGINAL ARTICLE

Open Access



Fusion of thermal and RGB images for automated deep learning based crack detection in civil infrastructure

Quincy G. Alexander^{1,4*} , Vedhus Hoskere^{2†}, Yasutaka Narazaki^{3†}, Andrew Maxwell^{1†} and Billie F. Spencer Jr⁴

Abstract

Research has been continually growing toward the development of image-based structural health monitoring tools that can leverage deep learning models to automate damage detection in civil infrastructure. However, these tools are typically based on RGB images, which work well under ideal lighting conditions, but often have degrading performance in poor and low-light scenes. On the other hand, thermal images, while lacking in crispness of details, do not show the same degradation of performance in changing lighting conditions. The potential to enhance automated damage detection by fusing RGB and thermal images together within a deep learning network has yet to be explored. In this paper, RGB and thermal images are fused in a ResNET-based semantic segmentation model for vision-based inspections. A convolutional neural network is then employed to automatically identify damage defects in concrete. The model uses a thermal and RGB encoder to combine the features detected from both spectrums to improve its performance of the model, and a single decoder to predict the classes. The results suggest that this RGB-thermal fusion network outperforms the RGB-only network in the detection of cracks using the Intersection Over Union (IOU) performance metric. The RGB-thermal fusion model not only detected damage at a higher performance rate, but it also performed much better in differentiating the types of damage.

Keywords: Infrared thermography, Structural health monitoring, Image fusion, Automated crack detection

1 Introduction

Infrastructure in the United States is in a growing state of disrepair, as the annual needs for repair/replacement funding continue to outpace the spending. The American Society of Civil Engineers (ASCE) estimated that the gap in infrastructure funding reached \$2 trillion dollars between 2016 and 2025 (ASCE, 2020). To combat this funding gap, infrastructure managers are moving towards an asset management-based model to maximize the impact of the limited funding. ASCE (2020) describes

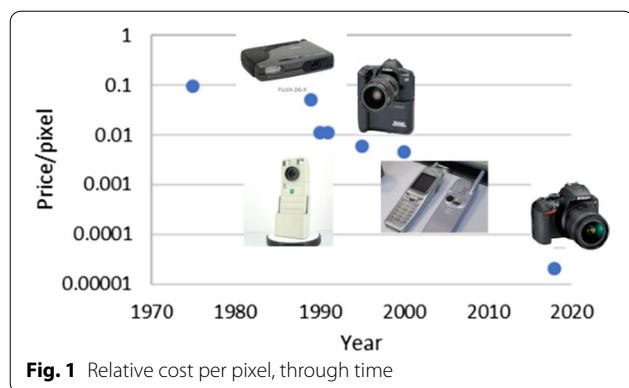
asset management as an effective tool for managing capital assets across various types of infrastructure to minimize the total cost of maintenance and operation in a dynamic and data-rich environment. Furthermore, it points out that the key to success of this proactive approach lies in the significant advances in the monitoring technologies, to prioritize needs and plan long-term strategies. As automated structural health monitoring (SHM) tools progress, their potential impact on asset management practices will continue to grow.

SHM tools employing deep learning have shown their efficacy in providing powerful data processing approaches for damage detection and structural condition assessments for a variety of infrastructure types (Bao & Li, 2020; Hess et al., 2015; Ye et al., 2019). Computer vision-based deep learning techniques, in particular, have seen a tremendous growth as computational power

[†]Vedhus Hoskere, Yasutaka Narazaki and Andrew Maxwell contributed equally to this work.

*Correspondence: quincy.g.alexander@erdc.dren.mil

¹ US Army Engineer Research and Development Center, 3909 Halls Ferry Rd, Vicksburg, MS 39180, USA
Full list of author information is available at the end of the article



continues to increase and the cost per pixel of data continues to decrease (Koch et al., 2015; Long et al., 2015). As noted by Barry Hendy of Kodak Digital Camera Technology has developed at an exponential improvement rate similar to Moore's law for microprocessors, with the pixels per dollar doubling annually when comparing cameras of similar feature level (Brynjolfsson & McAfee, 2014; Commons, 2015; Lucas, 2012). This trend across a range of camera types is illustrated in Fig. 1.

Leveraging the reduced technology cost and the increased computational power, there has been an exponential increase in the creation and deployment of computer vision-based inspection and infrastructure monitoring methods. For example, Hoskere et al. (2020) proposed the use of semantic segmentation for vision-based structural inspection, where a fully convolutional neural network was used to automatically identify multiple damage defects, like cracks, corrosion, spalling, etc., on certain structures of interest to the U.S. Army Corps of Engineers, such as lock gates and bridges. A detailed review of the current research, key challenges, and ongoing work related to computer vision-based tools for the inspection of infrastructure can be found by Spencer et al. (2019). Dong and Catbas (2020), and Avci et al. (2021) also provided a comprehensive summary of computer vision-based SHM tools and techniques at both local and global levels, including the factors that can affect accuracies.

Infrared thermography (IRT) is the science of detecting infrared energy emitted by an object, converting it into an apparent temperature, and displaying the result as an image (Fluke, 2021). Infrared (IR) wavelengths are longer than those found in the visible spectrum and are not visible with the human eye. IRT can often provide structural anomaly information that is not identifiable in the visible spectrum; indeed, the power of IRT for non-destructive evaluation (NDE) has been well documented (Avdelidis & Moropoulou, 2004; Hess et al., 2015;

ASTM International, 2013). Passive IRT, specifically, is commonly used in the NDE of large civil infrastructure, where artificially heating the specimen is not always feasible (Avdelidis & Moropoulou, 2004). For passive IRT, solar radiation serves as the heat source, and studies have been performed to understand the optimum times to capture images with the highest thermal contrast and that diurnal temperature variations can be adequate to support the use of IRT to detect defects in fully shaded regions (Alexander et al., 2019; Washer et al., 2013).

Visible (RGB) and thermal images each have their own strengths and weaknesses. For example, visible images can provide a clear representation of the scene at high resolutions and low cost per pixel, but light sources can affect the image quality. In contrast, thermal images are much more robust to variations in lighting conditions and can provide sub-surface information, but generally suffer from comparatively low resolution and weak crispness. Several manufactures have produced cameras that can capture both visible and thermal images simultaneously, thus providing more contextual information for the thermal images.

Research has been performed in various fields to determine how the two image types (RGB and thermal) can be combined to enhance their respective advantages, while neutralizing their disadvantages. For example, the automotive industry is particularly interested in the fusion of RGB and thermal images for object detection in autonomous vehicles, for which low light conditions and strong glares are common challenges (Sun et al., 2019). Liu et al. (2016) showed an 11% improvement over the baseline in pedestrian detection by using their proposed fusion model. Shivakumar et al. (2019) proposed an RGB-thermal fusion technique in the DARPA Subterranean Challenge to identify four different object classes (person, drill, backpack, and fire-extinguisher). An et al. (2018) used an image matching technique for crack detection, which can compare areas on thermal and visible spectrum images to determine to identify matching cracks, thus reducing the false positives. For each of the referenced applications, the fusion of thermal and RGB images performed better than adoption of only RGB images.

While the existing researchers have previously investigated the use of both visible and thermal images for damage detection in civil infrastructure, those studies are typically performed on simple laboratory test specimens that lacked the type of visual complexity that is present in the field. In addition, these studies have leveraged active IRT, which requires an artificial heating element to highlight the flaws for detection. Moreover, they are based on object detection, simply drawing a box around the identified cracks without considering the pixel-level accuracy

of the prediction, which is nevertheless crucial for damage severity assessment.

This paper proposes to fuse features from visible and thermal spectrum images to develop a robust automated damage detection strategy for in-service civil infrastructure. The novelty of the research effort is the quantification of the benefits of fusing thermal features within the neural network for a semantic segmentation model, with class predictions determined per pixel. To achieve the goal, a curated dataset of both damaged and undamaged in-service infrastructure is developed, with each visible-spectrum image having a corresponding thermal image as well as hand-labelled ground truth images for semantic segmentation. The thermal images are collected in the field using passive IRT technology, with a low-cost thermal imager that connects to a mobile device. This set of hardware better aligns with the tools that would commonly be utilized by inspectors in the field. Additionally, images of concrete joints are included in the dataset to represent crack-like visual patterns that are actually not associated with any damage. The model, therefore, not only has to detect the cracks, but also is able to differentiate between visually similar classes. The performance is then measured by the predictions of each pixel. This research demonstrates that the fusion of RGB and thermal images improves the performance of the model over the RGB only model in properly predicting cracks, as well as in differentiating between cracks and comparable features.

2 Methodologies

The dataset collected and used for training, validating and testing for this effort contains images of certain in-service infrastructure taken in the field. This section describes the curation of the dataset, with aligned RGB and thermal images of the damaged infrastructure and corresponding labelled ground truths for use in a semantic segmentation algorithm.

2.1 Data collection

The FLIR One Pro Gen 3 (Flir One) thermal camera was selected to collect data, due to its balance between a low price point and relatively high thermal resolution (Alexander & Lunderman, 2021). The FLIR One unit can capture thermal and RGB images simultaneously. When this camera is connected to a mobile device, the FLIR One mobile app is used as a viewfinder and to control the operations of the camera. Thermal images are captured at a resolution of 160×120 pixels for storage on the device and then can be decompressed to a resolution of 640×480 pixels when uploaded to a personal computer by using the FLIR software. The visible spectrum images have a resolution of 1440×1080 pixels. A preliminary

study about the reliability of the unit revealed that the camera's performance was found to be within the specification for temperatures ranging from 0°F to 120°F (Alexander & Lunderman, 2021).

The RGB-thermal pairs of images are annotated for semantic segmentation as part of this research effort. Annotating images for semantic segmentation is an arduous task, as each pixel in the image must be assigned a label. While the annotation of regular shapes, like polygons, can be completed with a few clicks, such amorphous shapes as cracks require meticulous attention to details. To develop a high-quality dataset for this study, annotations are conducted using InstaDam (Hoskere et al., 2021), an open-source software for damage annotation developed by the University of Illinois at Urbana-Champaign. The dataset collected and used for training, validating and testing for this effort contains images of in-service infrastructure taken in the field.

2.2 Image alignment

Fusion of images requires that the images are properly aligned. To match two images, Rao et al. (2014) outlined a normalized cross-correlation (NCC) approach, which works well when there is good structure within the two images. In this approach, one image is held fixed, while the position of the second image is moved pixel-by-pixel, with the quality of the match calculated at each position. The quality of the match is quantified through a coefficient of correlation, a value ranging from 0 to 1, where 1 indicates a perfect match. The position with the highest correlation coefficient should provide the best alignment between the two images.

As shown in Fig. 2, the native resolution of the thermal images is lower than that of the visible image; therefore, the thermal image should be scaled up to align the thermal scenes with the corresponding visible image. The thermal image is appropriately scaled, and then the NCC method is applied to the scaled thermal image to locate the position of the maximum correlation coefficient. The image size is made consistent with the RGB image by zero-padding. To qualitatively validate the accuracy of this approach, the two images are blended. Fig. 3 shows: (a) the RGB image, (b) the padded thermal image, and (c) the RGB-padded thermal blended images for one example scene. Finally, this approach is applied to all the image pairs in the dataset, using the iron palette (e.g., Fig. 3c) and greyscale palette used for the thermal images. However, the NCC approach is not effective for all image pairs and specifically works poorly for the images with low thermal definition. Therefore, some images have to be manually realigned.

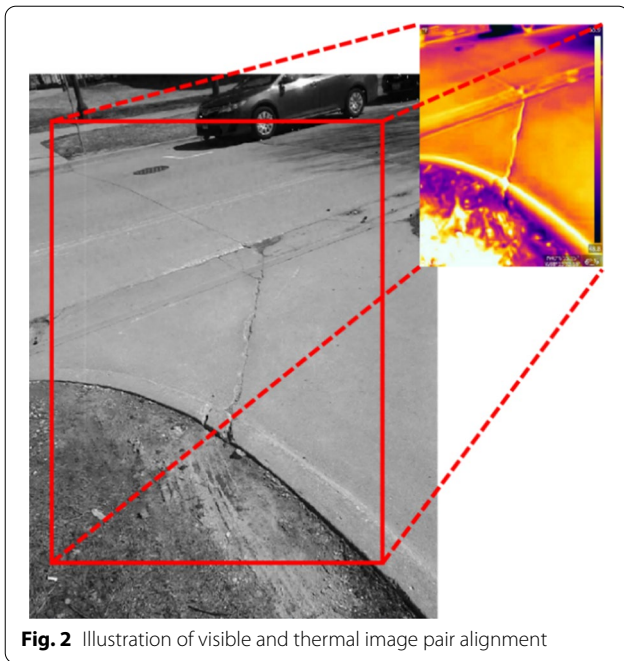


Fig. 2 Illustration of visible and thermal image pair alignment

2.3 Data preparation and annotation

Each image in the dataset has its corresponding pixel-wise annotations. Seventy-five of the images are selected with at least one crack and well-defined thermal contrast. The frequency of occurrences of different labelled classes is provided in Table 1. While the focus of this study is on crack detection, additional labels are included for spalling and vegetation growth. All the images are cropped to the size and position corresponding to the thermal images by removing the padded borders. The datasets generated during the current study are available from the corresponding authors on reasonable requests.

Table 1 Class label overview and weight

Class	Description	Pixels	Class probability	Weight
0	No label	242,470,928	0.9868	1.4357
1	Crack	2,252,136	0.0092	34.7848
2	Joint	998,536	0.0041	42.0544
3	Spalling	1,360,938	0.0055	39.654
4	Vegetation	2,678,104	0.0055	39.6544

2.4 Network architecture

The RTFNet network proposed by Sun et al. (2019) is used as the foundation for analysis in this study. The RTFNet architecture consists of an RGB encoder, a parallel thermal encoder, and a single decoder followed by the pixel-wise classification prediction. The encoder produces low-resolution feature maps for the RGB image and the thermal image, and the decoder up-samples the features to develop dense feature maps (Yasrab et al., 2017). The features acquired from each layer within the thermal decoder are mapped to the corresponding layer within the RGB encoder, as part of the fusion process. This network is illustrated in Fig. 4. The encoder is based on the Residual Network (ResNet) architecture, which has certain variants based on the number of layers. The ResNet-18 model with 18 neural network layers is used in this study.

Within the network, the classes are weighted based on the pixel distribution, according to the class weighting methodology outlined by Paszke et al. (2016). The class weighting formula is provided in Eq. 1. And the results are shown in Table 1.

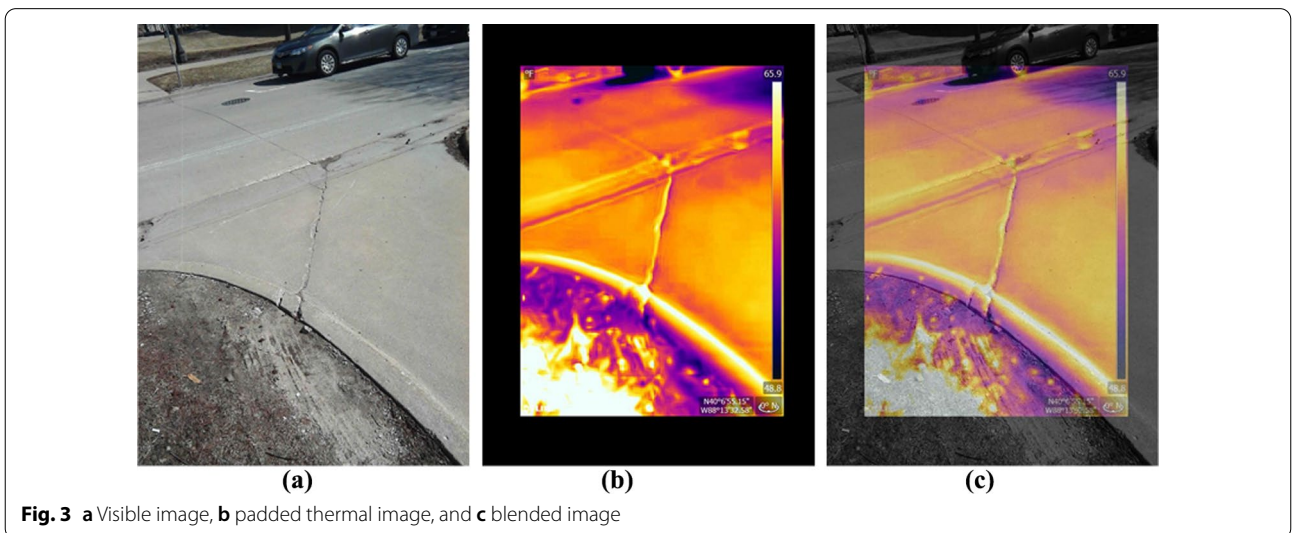
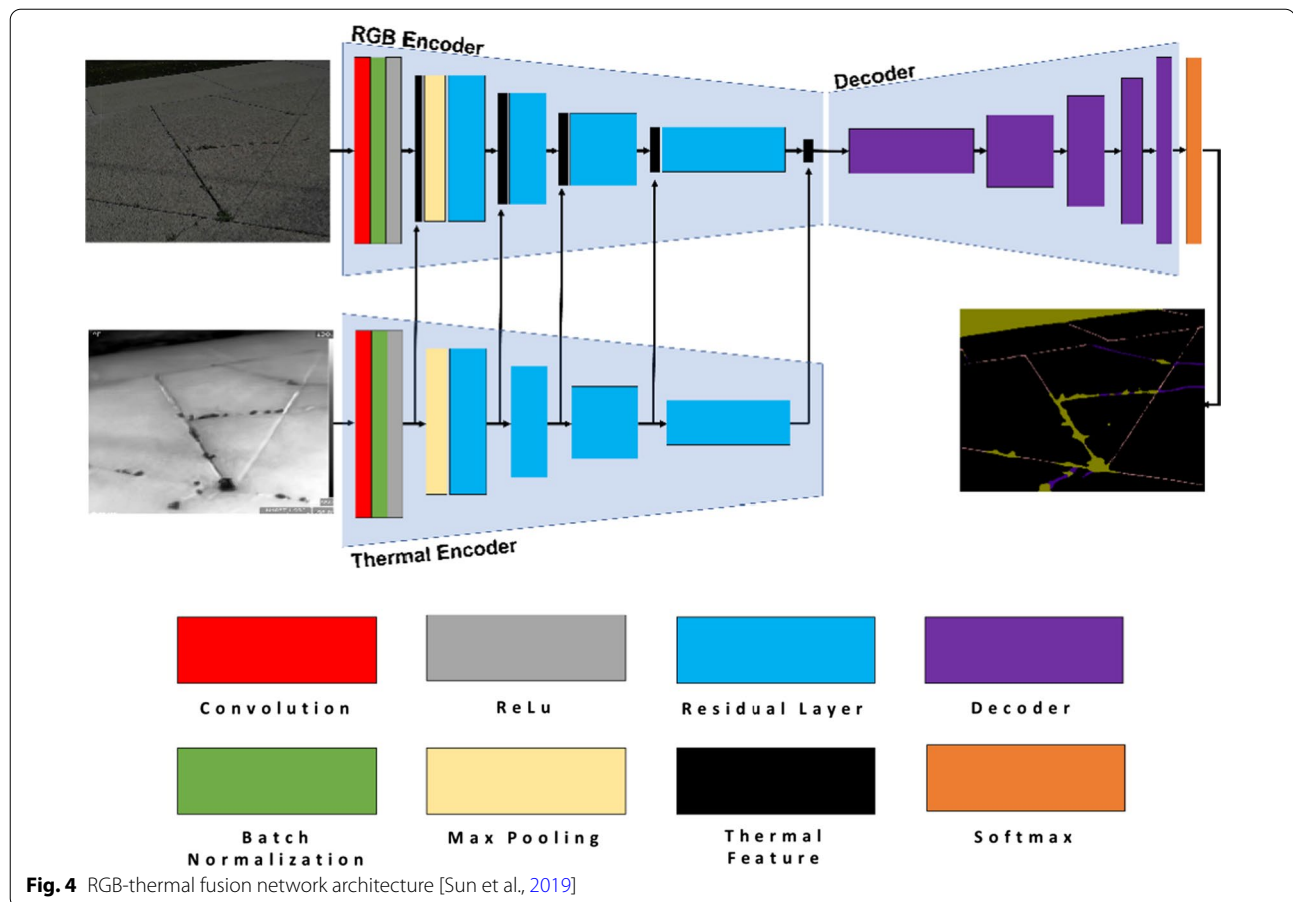


Fig. 3 a Visible image, b padded thermal image, and c blended image



$$\text{Weight} = \frac{1}{\ln(c + \text{class_probability})}, \quad (1)$$

where c is the Paszke Method Coefficient (1.02), and class_probability indicates the Ratio of the pixels of an individual class to the total number of pixels in the dataset.

Image augmentation schemes are applied to improve the training results. First, RGB images are duplicated, and the brightness of the matching images is reduced uniformly to simulate a low-light environment. The corresponding label images and thermal images are not modified. This augmentation method would double the size of the dataset, which is then randomly split by 80/10/10 for training/validating/testing, respectively. When the model is run for training, further data augmentations is applied: random flip, random noise, random brightness change, and random cropping.

3 Results

The following four scenarios are studied as part of the effort to quantify the value, including the thermal data:

- (1) Fusing the RGB and thermal images (RGBT). The greyscale version of the thermal images is used in the analysis.
- (2) Fusing the RGB images with a blank image (RGBB). This scenario represents the condition where only RGB data is available in an architecture that includes an empty (white) thermal input in the encoder.
- (3) Removing the thermal encoder from the architecture and analysing the RGB images only (RGB).
- (4) Removing the RGB encoder from the architecture and analysing the thermal images only (T).

3.1 Model performance evaluation

Both the RGBB and RGB models are included in the analysis to validate the process, as the RGB-blank pair should perform similarly to the RGB-only model. The performance of these four scenarios is measured in terms of Intersection over Union (IOU), which is one of the most common performance metrics used for semantic segmentation. At the pixel level, IOU

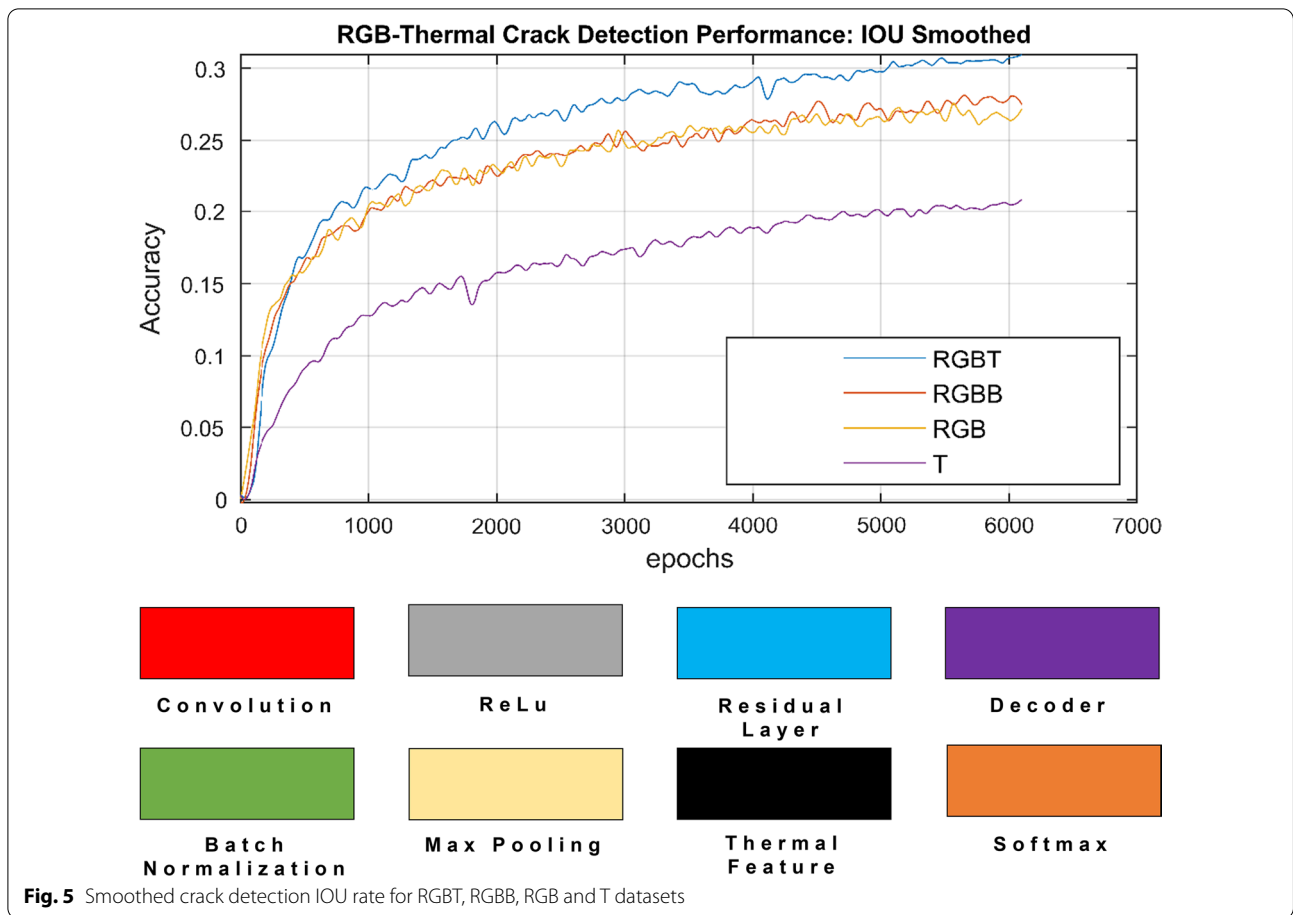


Table 2 IOU performance summary at 6000 epochs

Scenario	IOU crack detection performance
RGBT	0.31
RGB/RGBB	0.27
T	0.20

Table 3 Runtime comparison

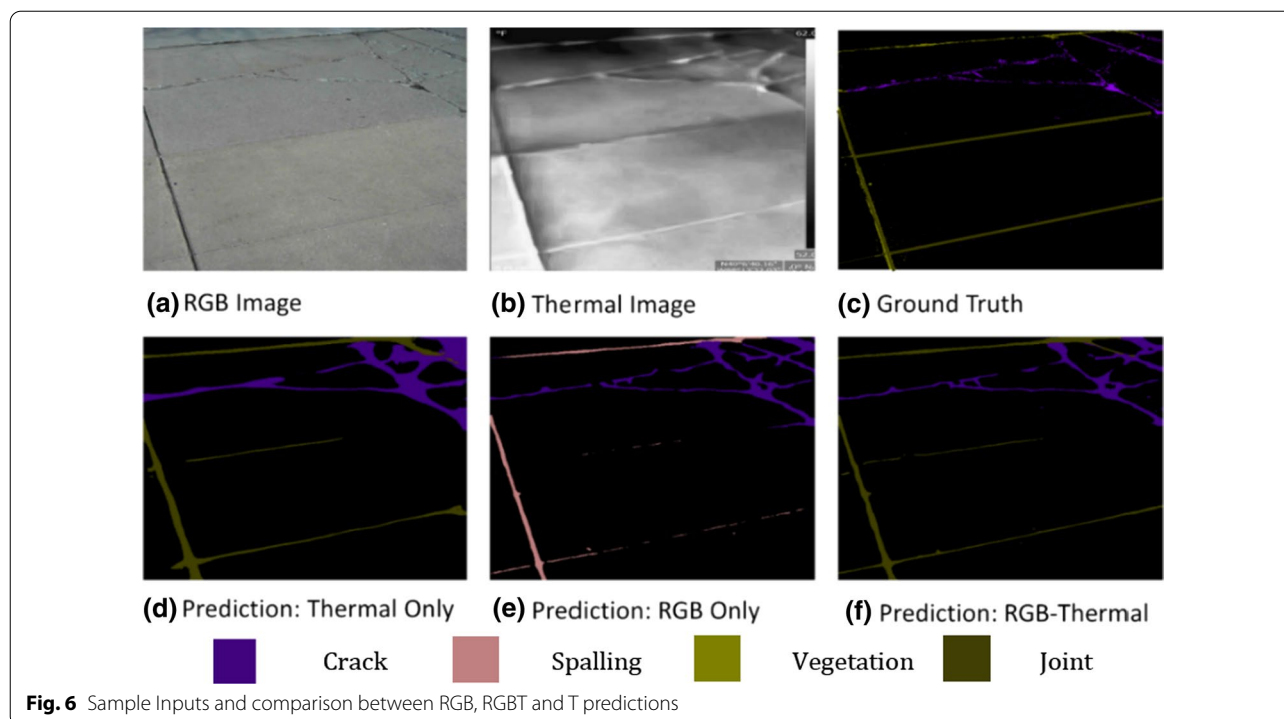
Model	Avg. seconds/epoch
Dual encoder	23.5
Single encoder	18.1

indicates the ratio of the correct class predictions to the sum of the correct and incorrect class predictions, as shown in Eq. 2. The performance for each scenario is shown in Fig. 5, after applying a Locally Weighted Scatterplot Smoothing (LOWESS) regression technique applied.

$$IOU = \frac{\text{Area of overlap}}{\text{Area of union}} = \frac{TP}{TP + FP + FN}, \quad (2)$$

where TP is the true positive (pixels), FP the false positive (pixels), and FN the false negative (pixels).

The results are shown in Fig. 5, and a summary of the performance at 6000 epochs is provided in Table 2. The results show that the fusion of RGB and thermal images outperforms RGB-only and T-only models, indicating that the network is able to leverage the additional information provided by the thermal images. By 6000 epochs, the performance of each scenario becomes stabilized relative to each other. The RGBT model outperforms the RGB-only model by approximately 15%. The RGBB network was trained to ensure that any performance improvement of the RGBT network over the RGB network was due to the additional information from the thermal images and not due to additional parameters in the network. The RGBB and RGB



accuracies align well with each other, signifying that the dual encoder systems with the second blank image perform similarly to an RGB-only image as expected. The results also show the ability of thermal images alone to provide an indication of cracks at approximately 74% of the rate of the RGB-only model, and 66% of the rate of the RGBT model. Thus, damage can be indicated in a scene with reduced impact of the lighting conditions, which supports the original hypothesis.

All the scenarios were run on the same device by using the same datasets for training, validating and testing. The run times (seconds/epoch) for the single-input and fused-input scenarios are summarized in Table 3, using a Predator PH317 with an Nvidia GeForce RTX 2070 GPU. Fusing the thermal data to the RGB data increases the run time by approximately 30 percent.

3.2 Qualitative results comparison

A sample of the inputs, labels, and predictions is provided in Fig. 6. As shown, some features, such as joints in the sidewalk, are challenging to be identified in the RGB image, but are prominent in the thermal image. The RGB-thermal pair performs well in predicting the joint locations and differentiating them from other classes, such as spalling. The thermal-only model's predictions are good for the class, but it lacks the sharpness in identifying the boundaries.

4 Further discussion

To illustrate the overall performance of the model, three specific conditions were evaluated. The first condition (Sample 1) displays a complex mix of classes; the second condition (Sample 2) represents certain visually similar materials with different thermal characteristics; and the third condition (Sample 3) represents low light conditions. These three conditions are highlighted in Fig. 7, with their performance compared in Tables 4, 5, 6. In addition to IOU, the recall rates are presented. In simple terms, recalls are used to measure the probability that a predicted class for a pixel is true. The equation is similar to IOU, except that the FN term is removed, so not recognizing a class is not as strongly penalized.

Sample 1, shown in the first column of Fig. 7, represents a scene with a complex mix of cracks and joints, as well as some vegetation in a well-lit scene. The results of this scenario indicate that the RGBT model is slightly better than the RGB model in identifying cracks. However, enhancements from fusing the thermal image are observed, as the RGBT model is much better at correctly differentiating between the classes. The RGB model misidentified joints as spalling, resulting in a joint recall and IOU score of 0. Vegetative growth was also misidentified as spalling. The T model had a comparable recall rate to that of the RGBT and RGB models, but with a slight reduction in IOU, as the thermal image lacked crispness to correctly maintain the boundaries of the classes.

	Sample 1	Sample 2	Sample 3
RGB input image			
Thermal input image			
Ground truth			
RGBT prediction			
RGB prediction			
T prediction			

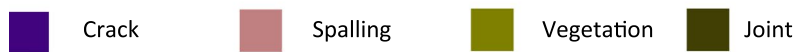


Fig. 7 Sample inputs and comparisons of output predictions

Table 4 Results of Sample 1

Model	RGBT	RGB	T
Crack IOU	0.36	0.32	0.29
Crack recall	0.97	0.98	0.97
Joint IOU	0.18	0	0.10
Joint recall	0.94	0	0.82

Table 5 Results of Sample 2

Model	RGBT	RGB	T
Crack IOU	0.28	0.25	0.24
Crack recall	0.95	0.95	0.94
Joint IOU	0.36	0	0.38
Joint recall	0.97	0	0.93

Table 6 Results of Sample 3

Model	RGBT	RGB	T
Crack IOU	0.25	0.22	0.20
Crack recall	0.83	0.68	0.62

Sample 2, shown in the second column of Fig. 7, represents a much simpler scene with a crack and a joint. In this sample, the material to the left of the joint is asphalt, and the material to the right of the joint is concrete. The two material types have different thermal properties. This difference can be seen in the thermal input image. All models correctly identified the crack, but the RGB model misidentified the joint as spalling, resulting in a joint recall and IOU score of 0, and overestimated the width.

Sample 3, shown in the third column of Fig. 7, represents a scene with only cracks, but in a low-light condition. The images were taken at night and included a pavement stripe for more visual complexity. In this scenario, the cracks were correctly identified by the RGBT and T models. But a portion of the crack patterns was misidentified as spalling by the RGB model.

In summary, these results show the significant potentials of the proposed RGBT approach in enhancing the efficiency and reliability of inspection of in-service civil infrastructures. Such inspection is required to identify structural damage, while robustly differentiating damage patterns from other similar patterns under various lighting conditions.

5 Conclusion

The purpose of this study was to quantify the value of fusing RGB and thermal images to improve a deep learning model for damage detection in large civil infrastructure. This is a novel approach for automated inspection of such infrastructure by leveraging the strengths of each image type, especially, where the features from each image type are fused at each layer of the deep-learning network. The RTFNet framework developed for autonomous vehicles was used as the foundation of this study. Images were collected by using a relatively inexpensive combined thermal and RGB camera, which was connected to a mobile device. Thermal-RGB image pairs were properly aligned, with annotations for semantic segmentation manually created for multiple classes, including cracks, joints, spalling and vegetation. Four scenarios were evaluated, including RGB-thermal fusion, RGB encoder only, RGB-fused with a blank image, and thermal image only. Each of the models was trained with over 6000 epochs, and using an 80/10/10 split for training, validating and testing. The results showed that the fusion of RGB and thermal spectrum images created a more robust model for the sample dataset, increasing the IOU value boosted by approximately 14% over the RGB-only model for crack detection, while providing more reliable class identification. The models trained with the thermal images alone delivered the lowest performance metrics. While the thermal-only model was generally capable of predicting the proper classes, the predictions lacked crispness and were often wider than the actual damage/joints. The predictions on the RGB images alone were not capable of consistently differentiating between the multiple class types, particularly in complex and low-light scenes. This study confirmed the hypothesis that fusion of RGB and thermal images can outperform the RGB-only and T-only models. Therefore, it is demonstrated that the network is able to leverage additional information provided by thermal images to provide a more robust model for inspection tasks of in-service civil infrastructures inspection tasks.

Author contributions

QA conceptualized the study, established the methodology, collected and curated the data set, carried out the study, and drafted the manuscript; VH participated in the design of the study, coordinated the data labeling, and helped review and edit the manuscript; YN participated in the design of the study, assisted in code modification, and helped review and edit the manuscript; AM established workflow for data curation and assisted in code development; BFS conceived of the study, participated in the design of the study, provided resources, and helped review and edit the manuscript. All authors read and approved the final manuscript.

Funding

Partial financial support was received from the U.S. Army Corps of Engineers Engineer Research and Development Center.

Declarations

Competing interests

The authors have no relevant non-financial interests to disclose. Billie F. Spencer Jr is an editorial board member for *AI in Civil Engineering* and was not involved in the editorial review, or the decision to publish, this article. All authors declare that there are no competing interests.

Author details

¹US Army Engineer Research and Development Center, 3909 Halls Ferry Rd, Vicksburg, MS 39180, USA. ²Department of Civil and Environmental Engineering, University of Houston, 4726 Calhoun Road, Houston, TX 77204, USA. ³Zhejiang University/University of Illinois at Urbana-Champaign Institute, Room C323, Engineering Building, 718 East Haizhou Road, Haining 314400, Zhejiang, China. ⁴Department of Civil and Environmental Engineering, University of Illinois-Urbana Champaign, 205 N Mathews Ave, Urbana, IL 61801, USA.

Received: 30 November 2021 Accepted: 18 March 2022

Published online: 18 August 2022

References

- Alexander, Q. G., & Lunderman, C. V. (2021). *Thermal camera reliability study: Flir One Pro*. US Army Engineer Research and Development Center.
- Alexander, Q. G., Hoskere, V., Spencer Jr, B. F., & Smith, D. M. (2019). Towards the application of image based monitoring of USACE Large Civil Infrastructure. International Workshop for Structural Health Monitoring. Palo Alto, CA.
- An, Y.-K., Jang, K.-Y., Kim, B., & Cho, S. (2018). Deep learning-based concrete crack detection using hybrid images. Sensors and Smart Structures Technologies for Civil, Mechanical, and Aerospace Systems 2018. Denver.
- ASCE. (2020). Changing the infrastructure equation: Using asset management to optimize investments. Retrieved March 20, 2021, from <http://preprod.asce.org/-/media/asce-images-and-files/advocacy/documents/changing-infrastructure-equation-report.pdf>
- ASTM International. (2013). D4788-03(2013) standard test method for detecting delaminations in bridge decks using infrared thermography. West Conshohocken: ASTM International.
- Avcı, O., Abdeliaber, O., Kiranyaz, S., Hussein, M., Gabboui, M., & Inman, D. J. (2021). A review of vibration-based damage detection in civil structures: From traditional methods to machine learning and deep learning applications. *Mechanical Systems and Signal Processing*, 147, 10707.
- Avdelidis, N. P., & Moropoulou, A. (2004). Applications of infrared thermography for the investigation of historic structures. *Journal of Cultural Heritage*, 5(1), 119–127.
- Bao, Y., & Li, H. (2020). Machine learning paradigm for structural health monitoring. *Structural Health Monitoring*, 20, 1353–13723.
- Brynjolfsson, E., & McAfee, A. (2014). *The second machine age: Work, progress, and prosperity in a time of brilliant technologies*. W.W. Norton & Company.
- Commons, W. (2015). File:Hendys Law.jpg. Retrieved April 18, 2020, from https://commons.wikimedia.org/wiki/File:Hendys_Law.jpg
- Dong, C.-Z., & Catbas, F. (2020). A review of computer vision-based structural health monitoring at local and global levels. *Structural Health Monitoring*, 20(2), 692–743.
- Fluke. (2021). What does infrared mean? Retrieved October 31, 2021, from <https://www.fluke.com/en-us/learn/blog/thermal-imaging/how-thermal-cameras-use-infrared-thermography>
- Hess, M., Vanoni, D., Petrovic, V., & Kuester, F. (2015). High-resolution thermal imaging methodology for non-destructive evaluation of historic structures. *Infrared Physics and Technology*, 73, 219–225.
- Hoskere, V., Fouad, A., Friedel, D., Yang, W., Tang, Y., Narazaki, Y., et al. (2021). InstaDam: Open-source platform for rapid semantic segmentation of structural damage. *Applied Sciences*, 11(2), 520.
- Hoskere, V., Narazaki, Y., Hoang, T. A., & Spencer, B. F., Jr. (2020). MaDnet: Multi task semantic segmentation of multiple types of structural materials and damage in images of civil infrastructure. *Journal of Civil Structural Health Monitoring*, 10, 757–773.
- Koch, C., Doycheva, K., Kasi, V., Akinci, B., & Fieguth, P. (2015). A review on computer vision based defect detection and condition assessment of concrete and asphalt civil infrastructure. *Advanced Engineering Informatics*, 29(2), 196–210.
- Liu, J., Zhang, S., Wang, S., & Metaxas, D. N. (2016). Multispectral deep neural networks for pedestrian detection. <https://arxiv.org/abs/1611.02644>.
- Long, J., Shelhamer, E., & Darrell, T. (2015). Fully convolutional networks for semantic segmentation. 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), (pp. 3431–3440). Boston.
- Lucas, H. C. (2012). *The search for survival: Lessons from disruptive technologies*. ABC-CLIO LLC.
- Paszke, A., Chaurasia, A., Kim, S., & Culurciello, E. (2016). ENet: Deep neural network architecture for real-time semantic. arXiv preprint [arXiv:1606.02147](https://arxiv.org/abs/1606.02147).
- Rao, Y., Prathapani, N., & Nagabhooshanam, E. (2014). Application of normalized cross correlation to image registration. *International Journal of Research in Engineering and Technology*, 3(5), 12–15.
- Yasrab, R., Gu, N., & Zhang, X. (2017). An encoder-decoder based convolution neural network (CNN) for future advanced driver assistance system (ADAS). *Applied Sciences*, 7(4).
- Shivakumar, S. S., Rodrigues, N., & Zhou, A. (2019). PST900: RGB-thermal calibration, dataset and segmentation network. Retrieved from <http://arxiv.org/abs/1909.10980>
- Spencer, B. F., Jr., Hoskere, V., & Narazaki, Y. (2019). Advances in computer vision-based civil infrastructure inspection and monitoring. *Engineering*, 5(2), 199–222.
- Sun, Y., Zuo, W., & Liu, M. (2019). RTFNet: RGB-thermal fusion network for semantic segmentation of urban scenes. *IEEE Robotics and Automation Letters*, 4(3), 2576–2583.
- Washer, G., Fenwick, R., Nelson, S., & Rumbayan, R. (2013). Guidelines for thermographic inspection of concrete bridge components in shaded conditions. *Transportation Research Record: Journal of the Transportation Research Board*, 2360(1), 13–20.
- Ye, X. W., Jin, R., & Yunc, C. B. (2019). A review on deep learning-based structural health monitoring of civil infrastructures. *Smart Structures and Systems*, 24(5), 567–575.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Submit your manuscript to a SpringerOpen[®] journal and benefit from:

- Convenient online submission
- Rigorous peer review
- Open access: articles freely available online
- High visibility within the field
- Retaining the copyright to your article

Submit your next manuscript at ► [springeropen.com](https://www.springeropen.com)