Smart Structures and Systems, *Vol. 24, No. 6 (2019) 723-732*
DOI: https://doi.org/10.12989/sss.2019.24.6.723

Development and testing of a composite system for bridge health monitoring utilising computer vision and deep learning

Darragh Lydon*[1], S.E. Taylor[1a], Myra Lydon[1b], Jesus Martinez del Rincon[2c] and David Hester[1d]

[1]School of Natural and Built Environment, Queen's University Belfast BT9 5AG, UK
[2]School of Electronics, Electrical Engineering and Computer Sciences, Queen's University Belfast BT9 5BN, UK

(Received July 12, 2019, Revised August 26, 2019, Accepted August 30, 2019)

Abstract. Globally road transport networks are subjected to continuous levels of stress from increasing loading and environmental effects. As the most popular mean of transport in the UK the condition of this civil infrastructure is a key indicator of economic growth and productivity. Structural Health Monitoring (SHM) systems can provide a valuable insight to the true condition of our aging infrastructure. In particular, monitoring of the displacement of a bridge structure under live loading can provide an accurate descriptor of bridge condition. In the past B-WIM systems have been used to collect traffic data and hence provide an indicator of bridge condition, however the use of such systems can be restricted by bridge type, assess issues and cost limitations. This research provides a non-contact low cost AI based solution for vehicle classification and associated bridge displacement using computer vision methods. Convolutional neural networks (CNNs) have been adapted to develop the QUBYOLO vehicle classification method from recorded traffic images. This vehicle classification was then accurately related to the corresponding bridge response obtained under live loading using non-contact methods. The successful identification of multiple vehicle types during field testing has shown that QUBYOLO is suitable for the fine-grained vehicle classification required to identify applied load to a bridge structure. The process of displacement analysis and vehicle classification for the purposes of load identification which was used in this research adds to the body of knowledge on the monitoring of existing bridge structures, particularly long span bridges, and establishes the significant potential of computer vision and Deep Learning to provide dependable results on the real response of our infrastructure to existing and potential increased loading.

Keywords: computer vision; multicamera; deep learning; structural health monitoring

1. Introduction

The civil infrastructure of a nation is a key indicator of economic growth and productivity (Romp and Haan 2007), possession of a reliable transport infrastructure facilitates production, tourism and many other commercial interests. In 2016, the United Kingdom (UK) government invested over £18.9 billion in infrastructure, with over 85% of this figure allocated to transport infrastructure (Office of National Statistics 2018). Facilitating over 90% of motorized passenger travel and 65% of domestic freight, the road network is the most popular means of transport in the UK. The road network is under continuous levels of stress from loading and environmental impacts whose effects can be detrimental to the integrity of the network. UK transport infrastructure is rated as second worst among the G7 countries (World Economic Forum 2018), and there is a bridge maintenance backlog valued at £6.7bn in 2019. (RAC Foundation 2019). In the UK, the budget for core bridge maintenance has been reduced by up to 40% in recent years (OECD 2016). This problem is extensible to most western countries. For instance, according to the 2017 Infrastructure report card, the corresponding figure in the USA is $123bn resulting in 188 million daily trips across structurally deficient bridges (ACSE 2017). This budgetary shortfall means that cost effective and accurate structural information on bridge condition is becoming increasingly important. According to literature (Graybeal *et al*. 2002, See 2012), the prevalent method for bridge monitoring continues to be visual inspections which can be highly subjective and differ depending on climatic conditions. Structural Health Monitoring (SHM) systems provide a valuable alternative to traditional inspections and overcome many of the previous limitations. SHM can provide an unbiased means of determining the true state of our ageing infrastructure. Sensor systems are used to monitor bridge deterioration and provide real information on the capacity of individual structures, hence extending the safe working life of bridges and improving safety. In particular, monitoring of the displacement of a structure under live loading provides valuable insight into the structural behaviour and can provide an accurate descriptor of bridge condition and provide validation for numerical assessments such as structural stress methods (Ye *et al*. 2015). Testing under live loading conditions also removes the requirement

*Corresponding author, Ph.D. Student
E-mail: D.Lydon@qub.ac.uk
[a] Professor
[b] Lecturer
[c] Lecturer
[d] Lecturer

http://www.techno-press.com/journals/sss&subpage=7
ISSN: 1738-1584 (Print), 1738-1991 (Online)

for bridge closure, an expensive undertaking which has knock on effects on other bridge structures as vehicles are diverted to alternative routes, increasing their loading. To monitor deterioration over time it is vital that the cause of displacement is also understood. Relating real time displacement along the span of a bridge to load type and location provides an opportunity to accurately identify localised damage within the structure. This research involves the use of Computer Vision methods for SHM. The basic principle involves using a camera to monitor the behaviour of a bridge as it experiences various effects – traffic load, varying temperature etc. Computer vision methods are used in this research as they are low cost, accurate and easily deploy able in the field. Previous research in this area has employed cameras to determine displacement, acceleration, natural frequency and response to temperature loads of bridges with varying span lengths. The bulk of the existing research is centred on single camera systems; this means that a trade-off between pixel accuracy and monitoring accuracy must be taken into consideration on bridges with a longer span. To reduce this factor, multi-point/multi-camera systems for displacement monitoring have been explored in the literature. Existing multi-camera methods require extensive cabling or impose range limitations on the system. This research will expand upon the current state of the art by successfully demonstrating the application of a wireless long range fully time synchronised application of displacement monitoring using Computer Vision. The findings presented in this paper show that displacement monitoring using computer vision methods can be used on long span bridges with the methodology of the system development presented in this paper. While accurate displacement information for abridge structure is useful, it is necessary to also identify the cause of this displacement through load identification. The established solutions for load identification have either used in situ instrumentation such as bridge weigh in motion (B-WIM) or coarse-grained video classification algorithms, this research will demonstrate the development of a fine-grained vehicle classification method.

The underlying challenge addressed by the research is that of obtaining accurate bridge response to correctly classified applied live load. Pairing these two inputs will allow the developed system to output statistical information to aid in the decision-making process for allocation of bridge maintenance resources.

2. Literature review

2.1 Review of multipoint displacement systems

In recent years, Computer Vision techniques have been applied to the field of SHM, allowing conventional Charge-coupled device (CCD) cameras to be used, as is discussed by Ye *et al*. (2016). One of the first studies carried out in this area was the work on the Humber Bridge using targets in conjunction with a template matching method performed by Stephen *et al*. (1993). In this work, a camera with zoom lens was focused on a high-contrast target that was installed at mid span of the Humber Bridge. Due to limits on storage capacity, a limited frame rate of 4.17 frames per second (FPS) was used for this testing which limits the information gathered from live loading. The results were compared to integrated acceleration values, only a visual comparison of the results was presented in the paper. While the lack of statistical comparison between the vision-based results and the traditional instrumentation is unfortunate, the results of the work were promising and were an early step in demonstrating the viability of using a camera to study bridge displacement. This experiment style, where a single camera was pointed at mid span of a bridge has been repeated several times in multiple studies (Ye *et al*. 2013, Zaurin and Catbas 2011, Fukuda, *et al.* 2010, Feng *et al*. 2015) with increasing accuracy as camera technology and image processing systems have developed over time. For the study of long span bridges, the use of multiple cameras has become increasingly prevalent. Multiple camera systems have a number of inherent advantages over single camera systems: multiple point measurements can be carried out on longer span bridges and the pixel: mm resolution can be greatly improved for multiple point measurements which reduces the measurement error. (Park *et al*. 2010) developed a partitioning method for studying the displacement of a high rise building where paired cameras would be used in series to monitor progressively higher levels of a tall building. Theoretically this system would result in a detailed analysis of building response, unfortunately in this study the authors do not explain or verify that their multiple cameras are accurately synchronised, so this application cannot be considered a viable multiple camera approach. A wired system for multi-point displacement using an optical flow method was demonstrated in laboratory work by Dong *et al*. (2018). Their results were very promising with a standard deviation between camera and potentiometer of 0.0154 mm. Their system for time synchronisation involved the cameras being cabled directly into the control unit, with time synchronisation handled by a USB timer. This type of centralised system can be time consuming to install and is limited to a centralized data server where data are collected and analysed that can become the main critical point of failure for the whole system. Additional multi-camera setups are trailed as shown in the literature by Malesa *et al*. (2016) and Malowany *et al*. (2017).

Fig. 1 Modified Action Camera with Zoom lens and synchronization hardware attached

A fully contactless time synchronized multi-camera system using modified action cameras (Fig. 1) for displacement (referred to below as QUBDisp) was developed and validated by the authors in work by Lydon *et al*. (2018). This system will be used in the field trials presented in Section 3 of this paper.

2.2 Review of composite systems for displacement measurement and analysis of bridge response

A hybrid sensor camera system for classifying vehicles was developed by Yan *et al*. (2008). In this paper, the authors laid out a system for grouping vehicles into seven different classes depending on the readings from electrical resistant strain gauges (ERSG) which were time synchronised with video images of the vehicles passing over a test bed. A neural network was developed to classify the vehicles based on the application of a Bayesian filter to the collected strain gauge readings. The main purpose of the video images was to establish the location of the vehicles on the deck. It was suggested that they could also be used to give gross weights of vehicles by assigning weights based on classes that were determined by an image based neural network. The classes were determined by annotating the database of images obtained from the cameras. In another research study by Fraser *et al.* (2010) a background subtraction method was used to identify the presence of a vehicle on the bridge. The background subtraction method is a technique where images of the empty bridge are used to provide a means of determining when a vehicle is travelling on the bridge. That is, images of the empty bridge are initially obtained, and then subtracted from images taken during a test. Any objects remaining after this subtraction are judged to be vehicles and subjected to further analysis for classification. Various techniques can be used for background subtraction, further reading can be found in the review by Piccardi (2004).

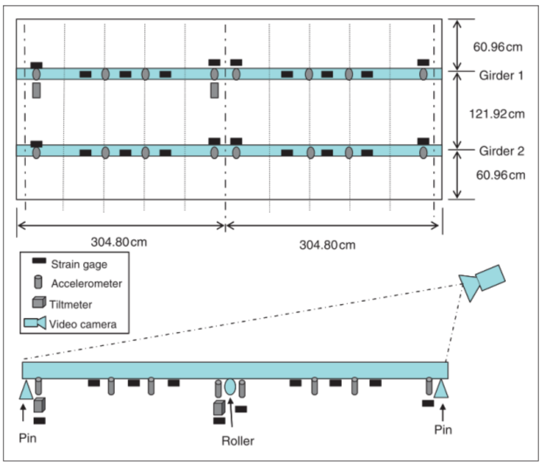

Fig. 2 Test Setup (Zaurin and Catbas 2011)

Images taken from a camera capturing images at 3 FPS were used with readings taken from paired with accelerometers scanning at 1000 Hz, foil strain gauges and thermocouples. The images from the camera were time stamped using LabVIEW to provide a means of synchronisation with the readings from the traditional instrumentation. The study did detect variations in natural frequencies in the range of 7-13%, without correlation of vehicle loads to bridge response it is difficult to determine the cause of this variance. Further laboratory-based testing of a camera-sensor system which utilised computer vision-based algorithms to determine the type of vehicle crossing a model bridge was carried out by Zaurin and Catbas (2011). The position of the vehicle at specified times was logged in order to build the Unit Influence Line (UIL) using the corresponding data obtained from 20 foil type strain gauges scanning at 1 kHz from the transducers placed on the underside of the model bridge as shown in Fig. 2.

The proposed system provided promising results, particularly in detecting changes in the bridge response based on various damage scenarios. However, this research is not transferable to field applications as it does not deal with an inherent issue of multiple sensor systems: time synchronisation. The data logger and USB camera in this study were linked to the same computer and could have been synchronised to the same time stamp. However, this particular approach is not feasible in the field as the USB camera did not have the required pixel resolution to detect deflection at the accuracy needed from typical distances in the field. Moreover, this vision-based system could only differentiate between three types of predetermined vehicle classes and a larger database would have created a more viable system for use in the field. This system was verified further by a field trial by Zaurin *et al.* (2015). In this work, video recordings were synchronised with strain readings from Hi-Tec weldable dynamic strain gauges that were attached to the main girders at the west and east bascule leaves of the Sunrise Bridge in Florida, USA. These readings were used to construct the Unit Influence Line (UIL) from numerous vehicle crossings on the bridge. The readings were synchronised using a Data Acquisition System (DAQ) which transmits data from each unit to a dedicated server using dedicated wireless network cards that are time locked through use of GPS timing receivers. Through the simulation of damage by adding and removing structural elements, the UIL created by the system can detect these changes, validating the presented method. The use of the background subtraction algorithm and requirement for strain gauges have significant limitations for field applications. Nevertheless, the underlying theory is a promising one and has provided the basis of the work carried out in this paper. A completely camera-based system for load identification and the corresponding displacement response measurement was developed by Ojio *et al*. (2016). The concept was to use two cameras which were time-synchronised, the first camera was used to measure sub-millimetre deflections on the underside of the bridge deck and the second camera to monitor passing vehicles on the bridge surface. The vehicle tracking camera was used a means of manually determining what vehicle type crossed

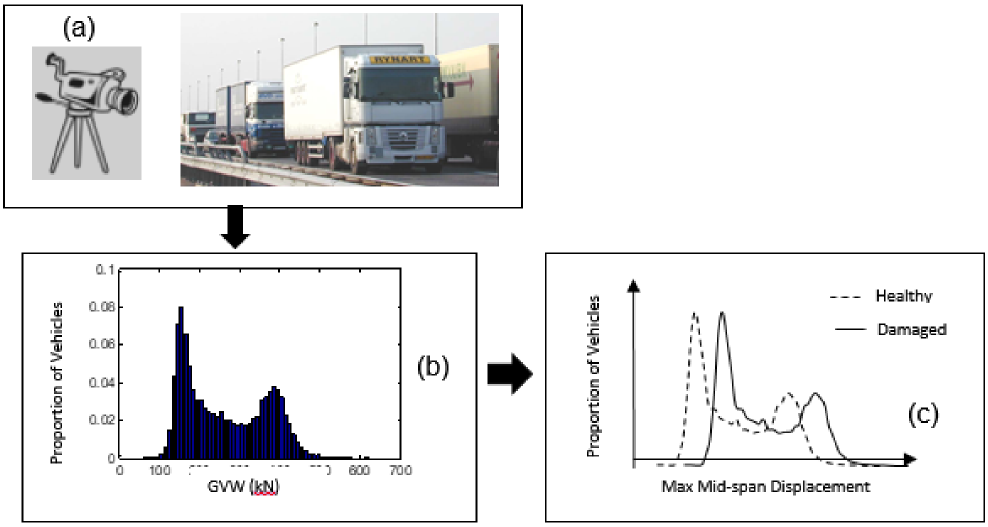

Fig. 3 Overview of proposed approach; (a) Camera identifies truck class/type on bridge and tracks it, (b) Histogram of gross vehicle weight for the identified class of truck and (c) Histogram of maximum mid-span displacement for that class of truck for the healthy and damaged cases

the bridge during a displacement event. There was no process to automatically detect or classify the vehicles crossing the bridge. This camera-based system provides a great advantage in situations where access to the underside of the bridge is restricted, thus making it very difficult to attach sensors particularly displacement which need to be independently supported. This research was also used as a basis for the research presented below. It was important the system could classify vehicles into types, as for certain classes of vehicle (e.g., 5-axle semi-trailers), the peak in the histogram of weights is highly repeatable, therefore it follows that the histogram of bridge deflections in response to that vehicle class, is repeatable. Once seasonal trends have been removed, any change in that most frequent deflection then indicates that the bridge's stiffness has changed, i.e., that there has been damage. This approach should be highly effective for long term monitoring.

2.3 Overview of the usage of convolutional neural networks for vehicle detection and classification

In recent years there has been significant research into the concept of deep learning, a subset of machine learning. Deep learning has been used in many applications, for example: driverless cars, translations, image colourisation and facial recognition with state-of-the-art results. The primary disadvantage to deep learning is that the training process requires a significant amount of data to obtain accurate results, with object classification tasks requiring thousands of images for accurate classification. Object classification and detection tasks using deep learning are typically performed using a Convolutional Neural Network (CNN). The concept of a CNN was inspired by a study by Hubel and Wiesel (1962) which discovered that different cells (or neurons) in the visual cortex were activated when the subject viewed basic shapes. This means that separate parts of the visual cortex were recognising different shapes, for example one cell could be responsible for determining if a shape is a straight line, another for curved lines and so on. When all these neurons are combined into a structure, the basis for visual recognition is produced. A CNN replicates this concept but does so using several layers of various types. Further reading on the concepts behind CNNs can be found in the literature (Cavaioni 2018). An implementation of GoogLeNet (Szegedy *et al.* 2015) for traffic detection was developed by Zhuo *et al.* (2017) an accuracy of up to 98% was achieved when classifying images from a data set created with the following classes: Large Bus, Car, Motorcycle, minibus, truck and van.

While the accuracy of this implementation is excellent, the requirement of having to crop and provide images to the CNN render it unsuitable for the purposes of this research. Additionally, it was desired to have a finer-grained differentiation between classes of truck, specifically by axle count, for the proposed implementation. Suhao *et al*. (2018) applied the Faster R-CNN (Ren *et al.* 2015) model using VGGNet (Simonyan and Zisserman 2015) for initial feature map generation to traffic detection over 3 classes: car, minibus and SUV. The mean Average Precision (mAP) (Worth 2010) of the classes detected ranged from 78%-85% which is acceptable, the lack of fine-grained traffic detection/classification is unsatisfactory for the requirements of our desired system. Faster R-CNN was also employed by Arinaldi *et al.* (2018) to detect and classify vehicles into either car, small van, bus, small truck or large truck categories. A mAP of 67% -69% was achieved over all classes based on a data set created from traffic videos. It was decided to use a similar image gathering method for this research as it enabled collection of 'real-life' data to be used for training the CNN, which can potentially improve accuracy of the system. Although these systems do show promising results in their desired application, they are not sufficient for replacing conventional methods for vehicle detections such as Bridge Weigh in Motion. Therefore, it was decided to implement a fine-grained system for vehicle detection and classification based on axle count.

3. Development and testing of proposed system for vehicle classification and displacement measurement

3.1 Development of vehicle classification method

The You Only Look Once (YOLO) method was chosen as the basis for creating a fine-grained system for vehicle classification (Redmon and Farhadi 2018). It was selected due to the excellent performance of YOLO compared to other CNN implementations. The underlying architecture for feature extraction in Yolo is Darknet-53 and is shown in Fig. 4.

YOLO uses anchor boxes to make predictions of object locations in an image. Anchor boxes give a general shape of objects to be detected and are calculated by using k-means clustering on the bounding boxes of objects used to train YOLO. YOLO then divides an image into an SxS grid. YOLO detects objects at multiple scales so S can vary depending on what scale of image has been passed through the input pipeline. The prediction confidence for object detection in each cell is then calculated, using Intersection over Union (IoU) between a predicted bounding box and the anchor boxes.

	Type	Filters	Size	Output
	Convolutional	32	3 × 3	256 × 256
	Convolutional	64	3 × 3 / 2	128 × 128
1×	Convolutional	32	1 × 1	
	Convolutional	64	3 × 3	
	Residual			128 × 128
	Convolutional	128	3 × 3 / 2	64 × 64
2×	Convolutional	64	1 × 1	
	Convolutional	128	3 × 3	
	Residual			64 × 64
	Convolutional	256	3 × 3 / 2	32 × 32
8×	Convolutional	128	1 × 1	
	Convolutional	256	3 × 3	
	Residual			32 × 32
	Convolutional	512	3 × 3 / 2	16 × 16
8×	Convolutional	256	1 × 1	
	Convolutional	512	3 × 3	
	Residual			16 × 16
	Convolutional	1024	3 × 3 / 2	8 × 8
4×	Convolutional	512	1 × 1	
	Convolutional	1024	3 × 3	
	Residual			8 × 8
	Avgpool		Global	
	Connected		1000	
	Softmax			

Fig. 4 Darknet-53 architecture (Redmon and Farhadi 2018)

Table 1 Class Breakdown of Image Dataset for QUBYOLO Training

Class	Image Count
Car	1,461
Bus	1,410
Van	1,170
Two-Axle Truck	1,559
Three-Axle Truck	1,537
Four-Axle Truck	1,788
Five-Axle Truck	1,276
Six- Axle Truck	831
Negative Images (Diverse Objects etc.)	7295

If the IoU values exceeds a user defined threshold YOLO determines that an object is present at that cell location in the scene. The centroid of the predicted bounding box is then converted to real world coordinates using a conversion ratio based on reference points in the scene. This was performed in this research by determining the location of, and the distance between, two pillars on the bridge used in the field trial. In order to create a valid implementation of YOLO for vehicle classification (hereafter referred to as QUBYOLO) an image database of different vehicle types had to be created. Numerous images of diverse objects that did not resemble the objects were also used to provide further data for comparison during training of QUBYOLO.

The images were collated from online databases and the data store was then increased using a process known as image augmentation. Image augmentation is a technique that slightly modifies an image passed to the process resulting in an image that, to a CNN, is distinct from the original version. This means that an image data set can be increased, allowing for more accurate training, without having to capture large amounts of new data. Prevalent augmentations include colour channel modification, horizontal and vertical flipping, translation, rotation and scaling. The total images per class is shown in Table 1.

A split of 60% training, 20% validation and 20% testing on each data set was performed for all image classification trials. Essentially, 60% of the gathered images were used to train QUBAN through the process of back propagation as explained above. 20% were used to validate training as it was being carried out; this process involves tuning the hyper parameters (e.g. learning rate, number of epochs, and the size of the image batch used to update the weights in a single pass) used in the training step by evaluating the predictive accuracy of an attempt by the network to classify a set of images. The network does not directly learn from these images; the results are only used as a basis for parameter modification. The final 20% of the image data set (which the trained CNN had never been seen before) is used to test the predictive accuracy of the trained network. The reason the images are split into 3 portions is that if the same images are used to both train and validate the network, the risk of over fitting increases. The separate testing portion is needed to approximate the real-world performance of the trained CNN on images it has not previously encountered. The average precision values for QUBYOLO at a threshold of 0.5 IoU between proposed bounding boxes and ground truth for testing images is shown in Table 2.

The mAP for the implementation of QUBYOLO was 73.55%. This high precision paired with the location of objects in the image scene resulted in QUBYOLO being deemed sufficiently accurate to be implemented in the evaluation of the full system for synchronised displacement measurement and load identification, which is detailed in the following section.

3.2 Evaluation of synchronised QUBYOLO & QUBDisp in field conditions

A field trial was carried out to investigate the applicability of QUBYOLO for fine-grained vehicle detection on traffic images with the time synchronised deflection readings from multiple cameras. The trial was carried out at Verners Bridge, Dungannon. Illustrated in Fig. 5, Verners Bridge is a 30 m span steel truss bridge situated on the Tamnamore road in Dungannon, Northern Ireland.

This road provides access to a busy industrial estate and is therefore frequently used by HGVs. Additionally, traffic on the bridge is controlled by a traffic light system which creates a single lane of traffic in one direction at any one point in time. This simplifies cases of multiple events on the bridge. Normal bridge behaviour was expected for this trial as the bridge was recently strengthened by Department for Infrastructure.

Table 2 AP for QUBYOLO

Class	Average Precision (%)
Car	86.77
Bus	80.49
Van	66.67
Two-Axle Truck	74.32
Three-Axle Truck	65.99
Four-Axle Truck	77.66
Five-Axle Truck	63.22
Six- Axle Truck	73.32

Fig. 5 Side elevation of Verners Bridge

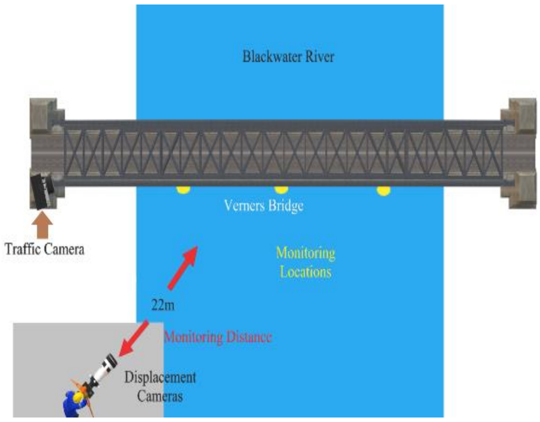

Fig. 6 Location of Cameras and Monitoring Location on Verners Bridge

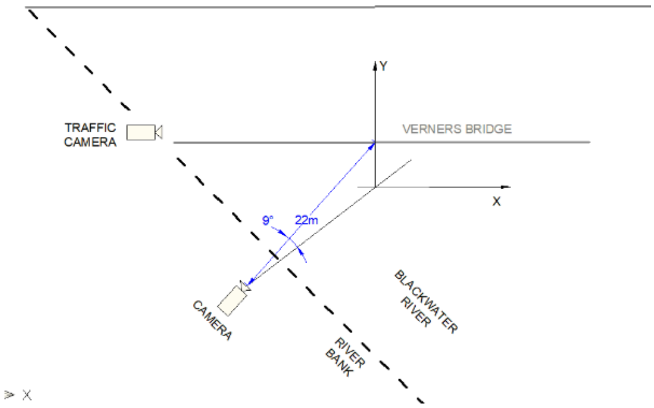

Fig. 7 Representation of Monitoring angle at Verners Bridge

Three GoPro cameras were used to track displacement on the bridge. The cameras were set up 13m from the bridge and set to the monitor the $\frac{1}{4}$, $\frac{1}{2}$ and $\frac{3}{4}$ span of the bridge. A 4th camera was deployed to record traffic images above the bridge deck. All 4 cameras were fitted with Syncbac accessories and set to record at 25fps. The location of each camera and its monitoring location on the bridge can be seen in Fig. 6 and a representation to show the monitoring angle can be seen in Fig. 7.

The $\frac{1}{4}$ span location was easily identifiable in the above deck image therefore this location was selected to extract images for QUBYOLO vehicle classification. The readings from each displacement camera were analysed using QUBDisp, the monitored displacements from 3 separate vehicle passes over the bridge is shown in Figs. 8-10.

Once the displacement for the truck events was obtained, it was possible to calculate an approximation of bridge response based on the applied load to the bridge structure. This was calculated for this trial by obtaining the displacement measurements when the first truck axle was at quarter span through a geometric transformation between the image coordinate (captured by the camera system) and the world coordinate of the bridge. An approximate curve of spatial displacement response was calculated using the

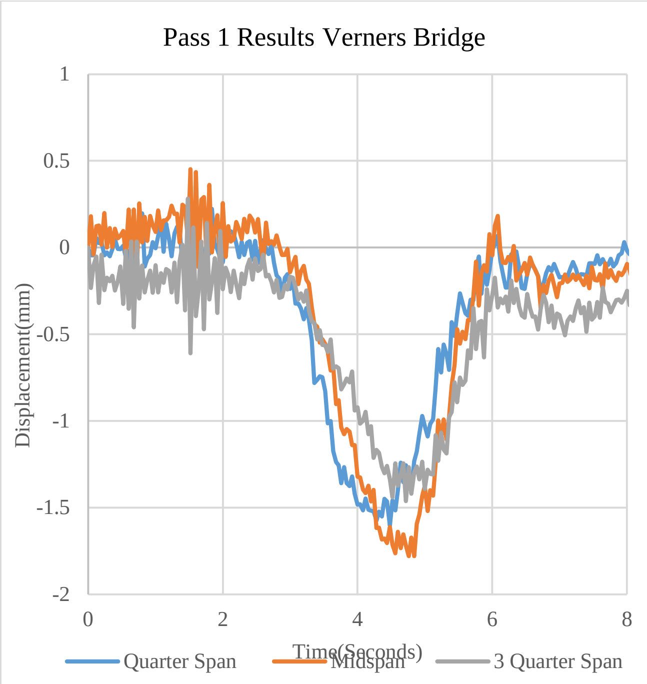

Fig. 8 Displacement from Pass 1 on Verners Bridge

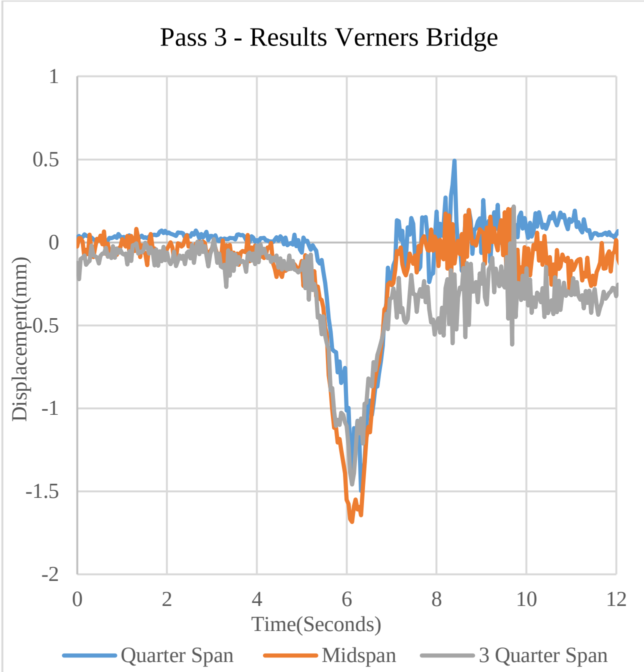

Fig. 10 Displacement from Pass 3 on Verners Bridge

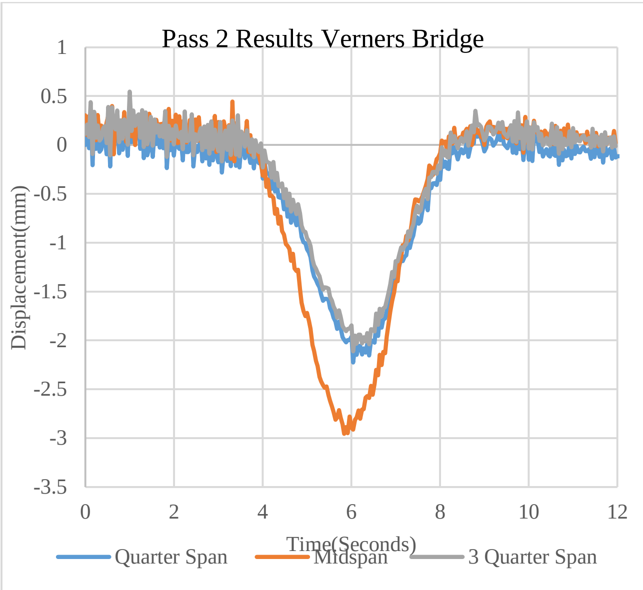

Fig. 9 Displacement from Pass 2 on Verners Bridge

curve fitting toolbox in MATLAB, it was assumed the bridge behaved like a simply supported structure. This displacement profile was used as a precursor to creation of a complete influence line that could be developed after access to a categorised weights database. The QUBYOLO implementation was then used to classify the images obtained from the traffic camera. QUBYOLO was not successful in all frames captured from the traffic camera, the precision (how accurate were the predictions) and recall (what proportion of all possible correct predictions were made) from the target class of vehicle in each pass are shown in Table 3.

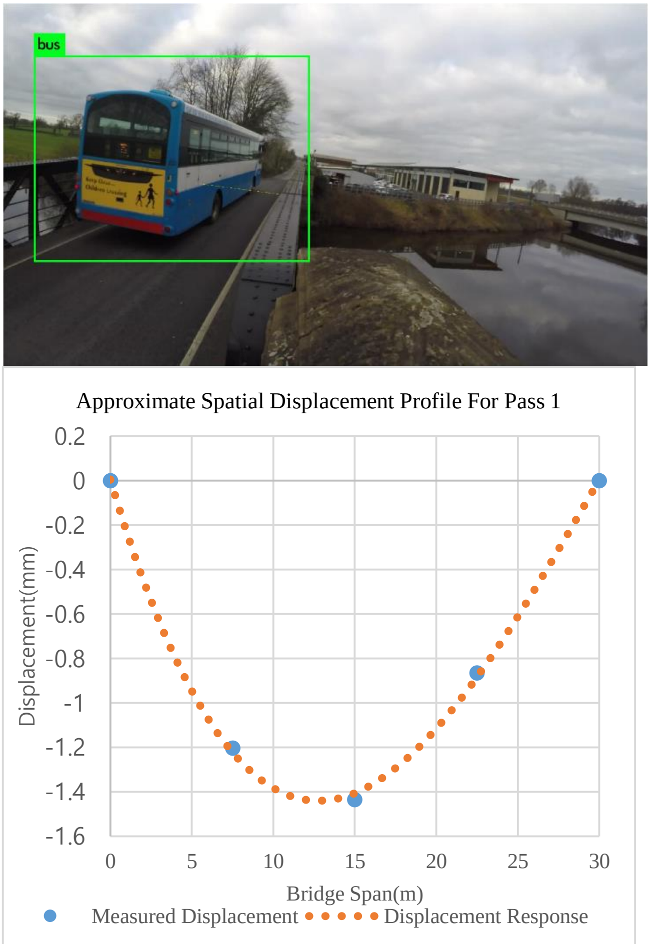

Fig. 11 QUBYOLO Prediction and Approximate Spatial Displacement Response for Pass 1 at Quarter Span

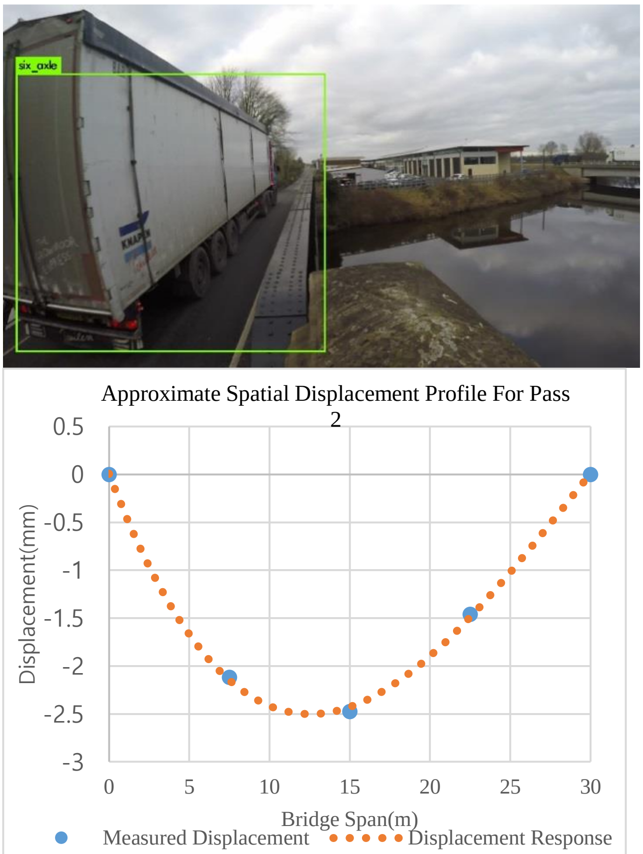

Fig. 12 QUBYOLO Prediction and Approximate Spatial Displacement Response for Pass 2 at Quarter Span

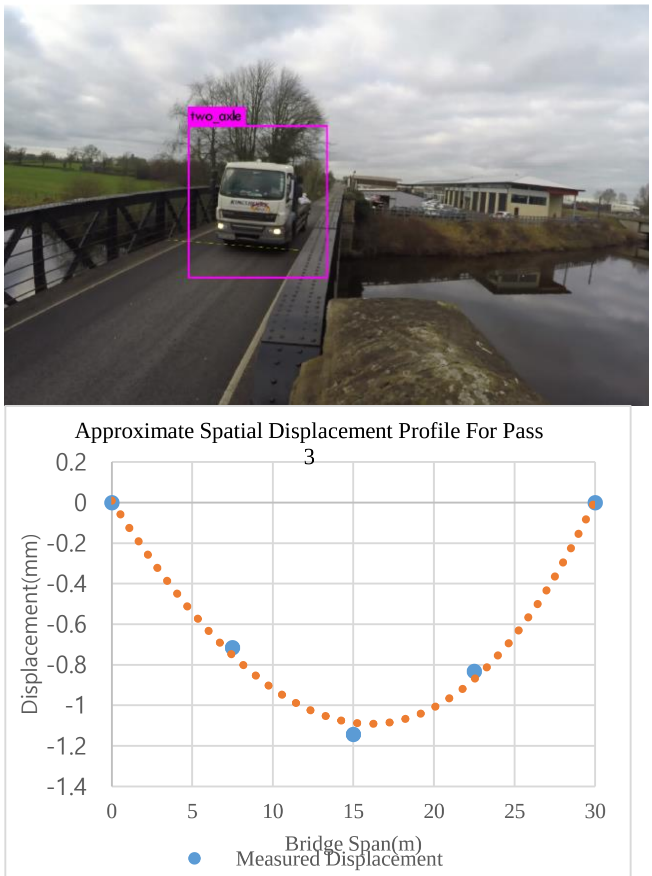

Fig. 13 QUBYOLO Prediction and Approximate Spatial Displacement Response for Pass 3 at Quarter Span

A decrease in performance was noted when the object was not centred in the field of view of the camera, most inaccurate predictions were made when this was the case. This was particularly problematic for the six axle HGV as not all axles were visible in the view of the traffic camera during image capture. Careful consideration needs to be taken regarding the camera position and orientation when setting up a monitoring location for a long-term implementation of this system. The use of images where vehicles were cropped and centred in the image to train QUBYOLO may also have led to this reduction in performance. Successful identifications from QUBYOLO as the front axle of each vehicle passed the quarter span monitoring location and the approximate curve of spatial displacement response from the bridge at this time are shown in Figs. 11 -13. The results obtained in this field trial confirm the successful identification of vehicle position/type in combination with monitoring of displacement of a bridge structure.

Table 3 Precision and Recall Results for Field Evaluation of QUBYOLO

Target Class	Precision (%)	Recall (%)
Bus (Pass 1)	78	58
Six Axle HGV (Pass 2)	45	60
Two Axle HGV (Pass 3)	63	50

4. Conclusions

This paper has presented a review of the current state-of-the-art in relation to vehicle detection for bridge monitoring systems using computer vision and adapted CNN. This review identified a critical need for a non-contact and non-destructive method of vehicle classification from recorded images which can be accurately related to the corresponding bridge response to live loading. Previous research presented AI vision-based traffic identification methods which were suitable for traffic studies with the potential for use in condition rating of bridges but with a need for improved accuracy. The development and testing of the QUBYOLO method has resulted in acceptable performance and greater applicability for the requirements of this research, that is, the algorithm has been proven capable of accurate detection and location of objects in an image scene. The field trial has also demonstrated the flexibility of the system as the camera used to record images of traffic was in a position and orientation that had not been used previously when capturing images for training. The successful identification of multiple vehicle types has shown that QUBYOLO is suitable for the fine-grained vehicle classification required to identify applied load to a bridge structure. No other research work has been found which has successfully classified HGV's into groups which are specific enough to facilitate load identification. The process of displacement analysis and vehicle classification for the purposes of load identification which was used in this research adds to the body of knowledge on the monitoring of existing bridge structures, particularly long span bridges, and establishes the significant potential of computer vision and Deep Learning to provide dependable results on the real response of our infrastructure to existing and potential increased loading. Future work in this area would involve the analysis of axle spacing of the captured vehicles, which could be used in combination with a vehicle weights database to create an influence line of bridge response to applied load. Access to a vehicle weights database categorised by axle count would facilitate calculation of approximate applied load for influence line creation. A damage detection method based on changes in the displacement profile for a specific weight classes over time can be implemented. This proposed system can act as an early warning system for bridge inspection and track change over time as a means of safety classification.

Acknowledgments

The authors wish to express their gratitude for the financial support received from Invest Northern Ireland, the USA National Science Foundation and Science Foundation Ireland towards this investigation under the US-Ireland Partnership Scheme. They also gratefully acknowledge Transport NI and the Department of Infrastructure NI for their support.

References

ACSE. (2017), "Report card for America's infrastructure", New York, NY, USA. URL: https://www.infrastructurereportcard.org/cat-item/bridges/

Arinaldi, A., Jaka, A.P. and Arlan, A.G. (2018), "Detection and classification of vehicles for traffic video analytics", *Pro. C. Sci.*, **144**(1), 259-268. DOI: 10.1016/J.PROCS.2018.10.527.

Cavaioni, M. (2018), "Deep learning series: Convolutional neural networks – machine learning bites – medium", Medium/Machine Learning. 2018. URL: https://medium.com/machine-learning-bites/deeplearning-series-convolutional-neural-networks-a9c2f2ee1524

Dong, C.Z., Celik, O. and Catbas, F.N. (2018), "Marker-free monitoring of the grandstand structures and modal identification using computer vision methods", *Struct. Health Monit.,* DOI: 10.1177/1475921718806895.

Feng, M.Q., Feng, D.M, Fukuda, Y. and Mizuto M. (2015), "Nontarget vision sensor for remote measurement of bridge dynamic response", *J. Bridge Eng.*, **20**(12). DOI: 10.1061/(ASCE)BE.1943-5592.0000747.

Michael, F., Ahmed, E., He, X.F. and Conte, J.P. (2010), "Sensor network for structural health monitoring of a highway bridge", *J. Comput. Civil Eng.*, **24**(1), 11-24. DOI: 10.1061/(ASCE)CP.1943-5487.0000005.

Fukuda, Y., Feng, M.Q. and Masanobu S. (2010), "Cost-effective vision-based system for monitoring dynamic response of civil engineering structures", *Struct. Control Hwalth. Monit.*, **17**(8), 918-936. DOI: 10.1002/stc.360.

Graybeal, B.A., Phares, B.M., Rolander, D.D., Mark, M. and Glenn, W. (2002), "Visual inspection of highway bridges", *J. Nondestruct. Eval.*, **21**(3), 67-83. DOI: 10.1023/A:1022508121821.

Hubel, D.H. and Wiesel T.N. (1962), "Receptive fields, binocular interaction and functional architecture in the cat's visual cortex", *J. Physiol.*, **160**(1), 106-154. DOI: 10.1113/jphysiol.1962.sp006837.

Lydon,D., Lydon, M., Del Rincon, J.M., Taylor, S.E., Robinson, D., O'Brien, E. and Catbas, F.N. (2018), "Development and field testing of a time-synchronized system for multi-point displacement calculation using low-cost wireless vision-based sensors", *IEEE Sens. J.*, **18**(23), 9744-9754. DOI: 10.1109/JSEN.2018.2853646.

Marcin, M., Krzysztof, M., Jakub, P., Malgorzata, K., Pawel, S., Artur, P., Tomasz, L. and Andrzej, Z. (2016), "Non-destructive testing of industrial structures with the use of multi-camera digital image correlation method", *Eng. Fail. Anal.*, **69**(11), 122-134. DOI: 10.1016/j.engfailanal.2016.02.002.

Krzysztof, M., Marcin, M., Tomasz, K. and Malgorzata, K. (2017), "Multi-camera digital image correlation method with distributed fields of view", *Opt. Laser. Eng.*, **98**(11), 198-204. DOI: 10.1016/J.OPTLASENG.2017.05.003.

Li, S.H., Lin, J.Z., Li, G.Q., Bai, T., Wang, H.Q. and Pang, Y. (2018), "Vehicle type detection based on deep learning in traffic scene", *Pro. C. Sci.*, **131**(2018), 564-572. DOI: 10.1016/J.PROCS.2018.04.281

OECD (2016), "Transport infrastructure investment and maintenance spending", Paris, France. URL: https://stats.oecd.org/

Office of National Statistics (2018), "Developing new statistics of infrastructure: August 2018", London, UK. URL: https://www.ons.gov.uk/economy/economicoutputandproductivity/productivitymeasures/articles/developingnewmeasuresofinfrastructureinvestment/august2018

Ojio, T., Carey, C.H., OBrien, E.J., Doherty, C. and Taylor, S.E. (2016), "Contactless bridge weigh-in-motion", *J. Bridge Eng.*, **21**(7), 4016032. DOI: 10.1061/(ASCE)BE.1943-5592.0000776.

Park, J.W., Lee, J.J., Jung, H.J. and Myung, H. (2010), "Vision-based displacement measurement method for high-rise building structures using partitioning approach", *NDT & E Int.*, **43**(7), 642-647. DOI: 10.1016/j.ndteint.2010.06.009.

Piccardi, M. (2004), "Background subtraction techniques: A review", *IEEE Inter. Confer. Sys.*, **4**, 3099-3104. DOI: 10.1109/ICSMC.2004.1400815.

RAC Foundation (2019), "Bridge maintenance backlog grows", London, UK. URL: https://www.racfoundation.org/media-centre/bridge-maintenance-backlog-grows

Redmon, J. and Farhadi, A. (2018), "YOLOv3: An incremental improvement." *arXiv:1804.02767,* DOI: 10.1109/CVPR.2017.690.

Ren, S.Q., He, K.M., Girshick, R. and Sun, J. (2015), "Faster R-CNN: Towards real-time object detection with region proposal networks", *IEEE T. Pattern Anal.*, **39**(6), 1137-1149. DOI: 10.1109/TPAMI.2016.2577031.

Ward, R. and Jakob D.H. (2007), "Public capital and economic growth: A critical survey", *Per. Der Wirts.*, **8**(1), 6-52. DOI: 10.1111/1468-2516.00242.

See, J.E. (2012), "Visual inspection: A review of the literature." *Off. Sci. Tech. Inform.*, 1-77. URL: https://pdfs.semanticscholar.org/5913/d3cbf8c389a8fdfe3fd5f2e880a0da03ba5a.pdf

Simonyan, K. and Andrew, Z. (2015), "Very deep convolutional networks for large-scale image recognition", *CoRR abs/1409.1556.*, DOI: arXiv:1409.1556

Stephen, G.A., Brownjohn, J.M.W. and Taylor, C.A. (1993). "Measurements of static and dynamic displacement from visual monitoring of the Humber Bridge", *Eng. Struct.*, **15**(3): 197-208. DOI: 10.1016/0141-0296(93)90054-8.

Szegedy, C., Liu, W., Jia, Y.P., Pierre, S., Scott, R., Dragomir, A., Dumitru, E., Vincent, V. and Andrew, R. (2015), "Going deeper with convolutions", *IEEE CVPR.*, DOI: 10.1109/CVPR.2015.7298594.

World Economic Forum (2018), "The global competitiveness report 2018", Cologny, Switzerland. URL: http://reports.weforum.org/global-competitiveness-report-2018/competitiveness-rankings/#series=ROADINF

Worth, D. (2010), "Introduction to modern information retrieval", *Australian Academic and Research Libraries.*, **41**(4), 305-306. DOI: 10.1080/00048623.2010.10721488.

Ye, X.W., Ni, Y.Q., Wai T.T., Wong, K.Y., Zhang, X.M. and Xu, F. (2013), "A vision-based system for dynamic displacement measurement of long-span bridges: Algorithm and verification", *Smart Struct. Syst.*, **12**(3-4), 363-379. https://doi.org/10.12989/sss.2013.12.3-4.363.

Ye, X.W., Yi, T.H., Wen, C. and Su, Y.H. (2015), "Reliability-based assessment of steel bridge deck using a mesh-insensitive structural stress method", *Smart. Struct. Syst.*, **16**(2), 367-382. https://doi.org/10.12989/sss.2015.16.2.367.

Ye, X.W., Dong, C.Z. and Liu, T. (2016), "Image-based structural dynamic displacement measurement using different multi-object tracking algorithms", *Smart Struct Syst.*, **17**(6), 935-956. https://doi.org/10.12989/sss.2016.17.6.935.

Zaurin, R. and Catbas, F.N. (2011), "Structural health monitoring using video stream, influence lines, and statistical analysis", *Struct. Health Monit.*, **10**(3), 309-332. DOI: 10.1177/1475921710373290.

Zaurin, R., Khuc, T. and Catbas, F.N. (2015), "Hybrid sensor-camera monitoring for damage detection: Case study of a real bridge", *J. Bridge Eng.*, **21**(6), 1-27. DOI: 10.1061/(ASCE)BE.1943.

Zaurin, R. and Catbas, F.N. (2011). "Structural health monitoring using video stream, influence lines, and statistical analysis", *Struct. Health Monit.*, **10**(3), 309-332. DOI: 10.1177/1475921710373290.

Zhuo, L., Jiang, L.Y., Zhu, Z.Q., Li, J.F., Zhang, J., and Long, H.X. (2017), "Vehicle classification for large-scale traffic surveillance videos using convolutional neural networks", *Mach. Vision. Appl.*, **28**(7), 793-802. DOI: 10.1007/s00138-017-0846-2.