

Transfer learning on convolutional activation feature as applied to a building quality assessment robot

*International Journal of Advanced
Robotic Systems*
May-June 2017: 1–12
© The Author(s) 2017
DOI: 10.1177/1729881417712620
journals.sagepub.com/home/arx



Lili Liu, Rui-Jun Yan, Varun Maruvanchery, Erdal Kayacan,
I-Ming Chen and Lee Kong Tiong

Abstract

We propose an automated postconstruction quality assessment robot system for crack, hollowness, and finishing defects in light of a need to speed up the inspection work, a more reliable inspection report, as well as an objective through fully automated inspection. Such an autonomous inspection system has a potential to cut labour cost significantly and achieve better accuracy. In the proposed system, a transfer learning network is employed for visual defect detection; a region proposal network is used for object region proposal, a deep learning network employed as feature extractor, and a linear classifier with supervised learning as object classifier; moreover, active learning of top-N ranking region of interest is undertaken for fine-tuning of the transfer learning on convolutional activation feature network. Extensive experiments are validated in a construction quality assessment system room and constructed test bed. The results are promising in a way that the novel proposed automated assessment method gives satisfactory results for crack, hollowness, and finishing defects assessment. To the best of our knowledge, this study is the first attempt to having an autonomous visual inspection system for postconstruction quality assessment of building sector. We believe the proposed system is going to help to pave the way towards fully autonomous postconstruction quality assessment systems in the future.

Keywords

Active transfer learning, deep learning, faster R-CNN, building quality assessment, mobile robot

Date received: 12 July 2016; accepted: 13 March 2017

Topic: Special Issue - Robotic Applications Based on Deep Learning
Topic Editor: Ming Liu

Introduction

Recent advances in robotics, sensor technologies, as well as exhaustive quality requirements of building sector inspire researchers to perform the postconstruction quality inspection of buildings automatically. An appropriate sensor system can induce the postconstruction quality assessment task over a large construction site semiautomatically. This study seeks to develop an automated construction quality assessment robot system (A-CONQUARS), as shown in Figure 1, which, to the best of the authors' knowledge, is the first attempt in this area. The proposed A-CONQUARS is able to inspect crack, hollowness, finishing defects, alignment, and evenness. It consists of a mobile robot for mapping and localization, a

thermal camera with heater for hollowness detection, a color camera for crack and finishing defects detection, and a 2-D laser scanner and an inclinometer for alignment and evenness detection. The whole system is integrated using LabVIEW with related National Instruments (NI) modules, and a tablet is used to generate A-CONQUARS report and remotely control the mobile robot. The A-CONQUARS report can also be

Robotics Research Center, Nanyang Technological University, Singapore

Corresponding author:

Erdal Kayacan, Nanyang Technological University, 639798, Singapore.
Email: erdal@ntu.edu.sg



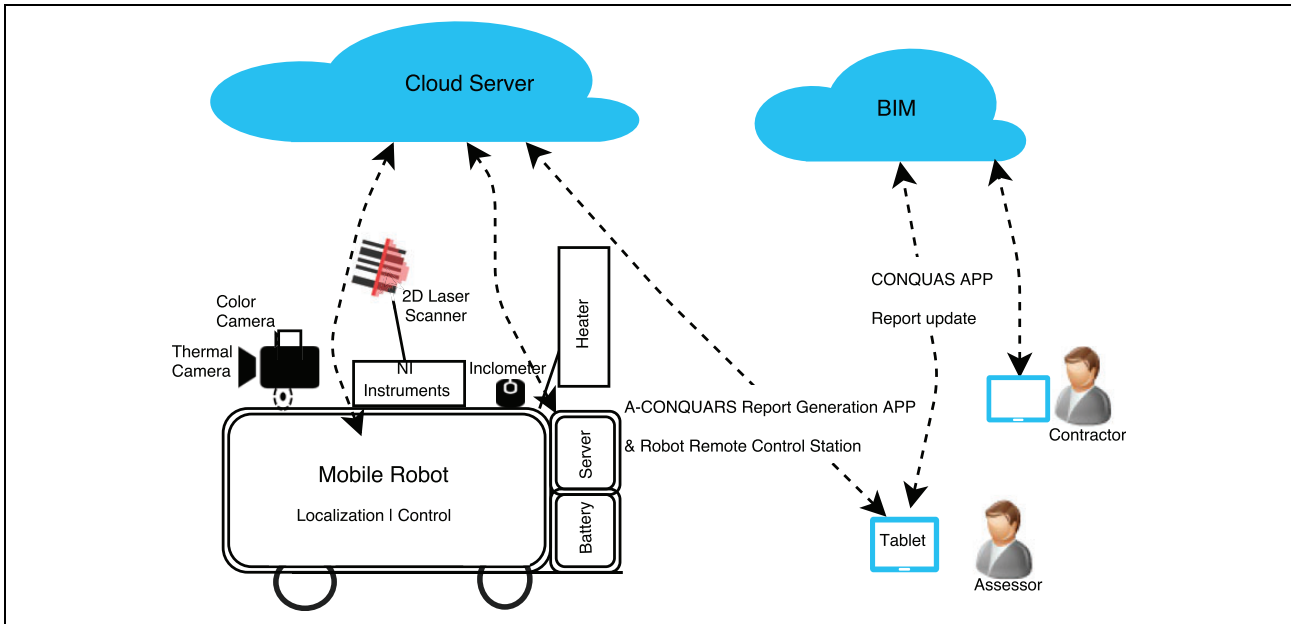


Figure 1. Automated construction quality assessment robot system.

used for the evaluation by assessor and provide a better understanding of defect location via the integration of building information modeling (BIM). Because of space considerations, this article focuses only on detection of crack, hollowness, and finishing defects using a proposed transfer learning (TL) network. All the three types of defects on the floors, walls, ceilings, doors, windows, and roofs can be automatically tracked using the proposed algorithm.

The contribution of this article is to propose a novel algorithm for real-time postconstruction visual quality assessment of buildings which is expected to improve the accuracy and efficiency of the assessment. This approach is named as transfer learning on convolutional activation features (TLCAFs), because it employs faster region-based convolutional neural networks (faster R-CNNs). The novelties of this investigation are as follows:

- (1) The TLCAF network is used to detect visual defects such as crack, finishing defects, and hollowness.
- (2) The proposed feature extraction approach achieves higher detection accuracy and higher discriminant power (DP) compared to some benchmark methods.^{1,2}
- (3) Assessment of hollowness underneath tiles of different materials is studied via thermal analyses. The thermal images captured can be efficiently analyzed by TLCAF to detect hollowness.
- (4) A robotic platform, illustrating how control, sensing, and actuation can be integrated to achieve an intelligent quality assessment system, is designed and presented.

The rest body of this article has the following sections: in “Introduction” section, the TL based on faster R-CNN

methodology is introduced; experiments and results are explained in “Experimental result” section. Finally, “Discussion and future work” section presents conclusions drawn from this study and some discussions on future work.

Related work

Review on approaches for hollowness assessment

Postconstruction quality assessment is an integrated part of any construction project which is currently carried out by manual inspection. In the standard assessment procedure, an assessor taps tiles or sweeps floor using a CONQUAS rod (a steel rod attached to a wooden stick) to identify the hollow-sounding tiles. Hollow space under tile can weaken the tile and cause breaking. For automatic inspection, the following techniques are used in literature review below. In recent 20 years, advanced nondestructive testing (NDT) technologies using radar, ultrasonic, and impact-echo have been used for the assessment of concrete and masonry structures. They are suitable for the detection and description of inhomogeneities in concrete and masonry structures at depths between 5 and 100 cm. In the near surface region between 0 and 10 cm, active infrared thermography (IRT) is preferred.³ IRT is reliable for detection of construction surface quality, since it is very sensitive to surface variations. Using IRT, unlike intact surfaces, a portion of surface having a defect has discontinuities in temperature distribution. Using active IRT, in order to enhance thermal contrast, surface is heated up using an external heat source. Maierhofer et al.³ used impulse thermography to detect the hollowness under the tiles. Similarly, long-pulse exposure IRT (LPT) in detecting surface flaws including cracks is

preferred.^{4,5} It is shown that the LPT is more appropriate for bond integrity detection, because it provides a fast and full-field measurement which is noncontact and easy for interpretation. In 2013, Brown and Hamilton⁶ heated the concrete samples using halogen lamps and used pulsed IRT to observe the defect area via transient heat stimulation. The thermogram is analyzed using normalization method on the acquired temperature versus time history for each pixel analyzed. Through the literature review above, it reveals that active infrared radiation thermography technology is mature and ready to be integrated into robotic systems for hollowness assessment of large complex composite structures.

Most thermal NDT processing algorithms use pixel-based temperature function. The efficiency of these algorithms can be evaluated by classification for defect recognition.^{7–10} For diagnosis of thermal defects, regions of interest (ROIs) are selected by feature descriptions. Various intelligent techniques, such as neural network (NN),^{11,12} support vector machine (SVM),^{13,14} and neuro-fuzzy algorithm,¹⁵ have been used for the classification. In literature, the simplest approach to distinguish hot/cold spot regions in the thermal image of a building is to use statistical methods and morphological image processing technique in conjunction with quantitative analyses on the inspection results.^{16,17}

Review on algorithms for crack detection

Manual visual assessment by human eyes is an effective approach for building sector quality assessment. However, it is subjective and depends on the experience and mental focus of the assessor which is prone to human assessment precision. The development of intelligent visual defect assessment approach can overcome the aforementioned shortcoming. In the study by Jahanshahi and Masri,² the crack segmentation parameters are adjusted in an adaptive manner based on depth parameters, and the crack features are extracted using linear discriminant analysis (LDA); a performance of an SVM, a nearest-neighbor classifier, and an NN are evaluated for classifying cracks from noncrack patterns. This research led to the development of an autonomous crack quantification method using the obtained crack map which can be applied for crack thickness detection. In the study by Nashat et al.,¹ pyramid SVM after Wilk's λ analysis is used for biscuit's crack inspection, and the approach demonstrates the effectiveness of hough-based features for crack detection. A crack inspection approach for bridge deck using a robotic crack assessment and mapping system is developed by Lim et al.,¹⁸ where cracks are detected using Laplacian of Gaussian algorithm,¹⁹ and the global crack map is obtained through camera calibration and robot localization. Automatic decision-making for defect recognition can be performed by establishing a signal threshold or defining defect features. Points of interest on the object can be extracted to

provide a feature description of the defects. The features need to be trained to get improvements, but the features themselves have no self-learning capability. NN can learn crack features including subtle variations in signal evolution, but each class of testing defects needs particular training of the network.

Tools for R-CNN-based detection

Recent improvements in region proposal methods and region-based convolutional NNs have contributed to object detection area.²⁰ Usually, region proposal network (RPN) is adopted as external modules that are independent of the detectors (e.g. selective search object detectors,²¹ R-CNN,²⁰ and fast R-CNN²²). Region proposal methods include sliding windows methods (e.g. objectness in windows²³ and edge boxes²⁴), grouping super-pixel methods (e.g. multiscale combinatorial grouping (MCG)²⁵, constrained parametric min-cut²⁶, and selective search²¹), metrics and in-depth analysis methods for Red Green Blue-Depth (RGB-D) data,²⁷ and regionlet for generic object detection.

Convolutional layers with deep networks for object detection can be applied to images of arbitrary size to yield proportionally sized feature map. R-CNN uses selective search's faster mode; it is slow since it goes through a ConvNet forward pass for each object proposal without sharing computation.²⁰ Spatial pyramid pooling in deep convolutional NNs (SPPnets)²⁸ is developed to speed up R-CNN by sharing computation for different convolutional feature maps. It accelerates R-CNN by 10 to 100 times in terms of test time. However, its training is a multistage pipeline. Compared to SPPnets, fast R-CNN network takes input as a whole image and a set of object proposals. It overcomes the disadvantages of R-CNN and SPPnet and is faster to train and test. Its training is single-stage by using a multitask loss, and it has higher detection quality (mean average precision).²² From region proposal module comparisons in Hosang et al.,²⁹ edge box, selective search, and MCG have good detection rate, and edge box shows the best trade-off between proposal quality and speed. However, the region proposal step still spends a relatively long time. Faster R-CNN merges RPN and fast R-CNN into a unified network by sharing the convolutional features with "attention" mechanisms, that is, RPN proposes the ROIs to fast R-CNN detector for defect recognition. Faster R-CNN³⁰ employs signal time-scaled SVD to increase feature speed and region-based bounding box scanning method to increase signal-to-noise ratio (SNR). It learns features all along real-time object detection with RPN, but it requires special training of the network. As a state of the art detection network, faster R-CNN reduces the running time and aims toward real-time object detection, and it fulfills intelligent building quality assessment condition. In addition, a new object detection method, named single shot multiBox detector (SSD)³¹ was released by Google

recently. It is the first deep network-based object detector and yields improvement in trade-off between speed and accuracy. SSD300 shows similar accuracy as faster R-CNN in 8 layers network, and SSD500 provides better performance than faster R-CNN in 16 layers network. This work uses deep learning models as feature extracting tool and carries out TLCAFs for visual defect detection of building sector.

Methodology

TLCAF framework

Deep learning usually requires millions of training data. In this study, only limited images are available for the tasks of visual defect detection; hence, a TL approach is explored. TL has the capability to recognize and apply knowledge learned in previous domains/tasks to novel domains/tasks,³² when source (previous) and target (novel) domains/tasks have some commonality (e.g. overlapping features).³³ Feature-based TL approaches learn the transformation by encoding application-specific knowledge (e.g. deep learning). Some higher level task/domain-specific features can help the target learning task even only a few labeled data are given.

Vision's deep learning's hierarchical composition includes pixels → edge → texon → motif → part → object .

Considering visual defects including crack, finishing defect, and hollowness has the commonality with source data including some low-level sudden change (gradient decent) of edge and/or texture. For higher level specific features of visual defect, feature-based TL is applied for defects detection. The TLCAF network is proposed and illustrated in Figure 2. Four steps for TLCAF are depicted below:

- (1) TLCAF from unlabeled data (faster R-CNN pre-trained detection model).
- (2) Represent visual defect by learned higher level features.
- (3) Train a model from the new representations of visual defect with corresponding labels.
- (4) Actively learn top-N ranking features for fine-tuning.

After the model is trained to represent visual defect features, the linear classifier (e.g. SVM/softmax) is switched to test model for defect prediction. A nonmaximum suppression layer is went through for calculating the defect score corresponding to the proposed bounding box (bbox). Finally, top-N ranked defects' bboxes are drawn on the detected images.

Significant improvement in object recognition rate can be achieved via fine-tuning³⁴; therefore, this is performed in step 4 for active TLCAF (A-TLCAF) learning.

TLCAF approach

Since the new data set of visual defect is small, faster R-CNN's network is transferred and used for feature extraction. Faster R-CNN is developed based on RPN and fast R-CNN detector. The RPN serves as the "attention"³⁵ mechanisms of the unified network by reducing proposals to top-N ranked proposal regions via objectness score; subsequently, RPN outputs these object regions to faster R-CNN for recognition and defects detection. For an input of an image of arbitrary size, RPN produces a set of rectangular region proposals, and each proposal has an objectness score.³⁰

Region proposals are parameterized relative to reference boxes. In faster R-CNN, the region is a generic term and only rectangular regions are considered. Objectness is commonly employed to quantify the probability of a set of object classes.^{21,24,36} Caffe deep learning framework is built and applied to eight layer Zeiler and Fergus model³⁷ in this work. RPN module has two sibling output layers which is referred to multitask loss mechanism in fast R-CNN: one is a discrete objectness probability distribution (per ROI), and the other is bbox regression offsets.

The cells in different convolutional layers act as local filters over the input space. Rectified linear unit (ReLU) is used as an activation function after each convolutional layer in place of the traditional sigmoid function, because it is not necessary for hidden neurons to have bounded values, and ReLU does not have gradient vanishing problem that exists for tanh and sigmoid functions. The normalization layer (i.e. local contrast normalization) for local response turns out to be useful when neurons with unbounded activations like ReLU are employed. It performs a kind of "lateral inhibition" by normalizing local input regions, because it is able to detect high-frequency features with a big neuron response, and meantime hamper responses that are uniformly large in a local neighborhood. It encourages "competition" for big activities among nearby neuron groups. For subsampling, max pooling is employed. The convolutional NNs are trained via back propagation and stochastic gradient descent. Convolutional proposal layer is shared by RPN and fast R-CNN. Another two sibling layers are the objectness score layer and the bbox regression layer of the proposals, and they are used to select top-N ranking region proposals. After the two sibling layers, data with the ROIs are passed to fast R-CNN object detection network, where the data are processed by one more time pooling to attain 4096 byte features, and two more fully connected layers with dropout mechanism to improve generalization. Dropout is a new technology to prevent NNs from overfitting by blocking complex coadaptations on training data. It provides high efficiency for model averaging with NNs. The 4096 byte/dimension features for each proposal after fully connected layer 7 (fc7) are extracted and built into a data set with labels, which is sent to linear SVM for training and testing. After that, the positive prediction result is combined with its bbox for

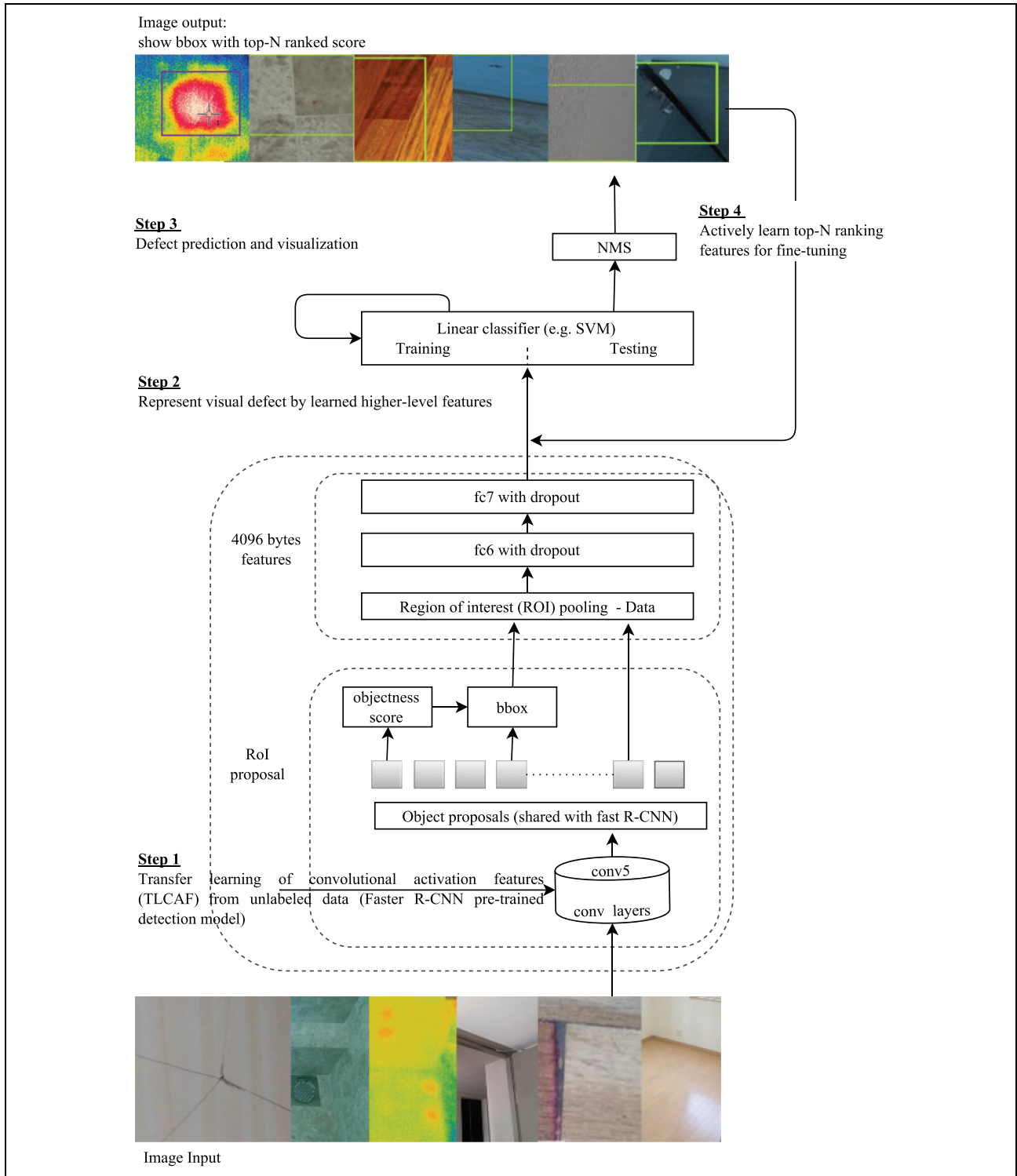


Figure 2. Active TLCAF network. TLCAF: transfer learning on convolutional activation feature.

defect detection. In addition, Zeiler and Fergus³⁷ show convolutional layers and fully connected layers, and both contribute to the decrease in classification error rates. That is the reason why features from the output of fc7 are extracted in this investigation. Meanwhile, the increase of filters in convolutional layers 3, 4, and 5 also yields improvement in

classification accuracy.³⁷ Furthermore, the increase in data set and/or further fine-tuning procedure for classification continuously improve accuracy rate for object prediction, rather than inheriting pretrained features.³⁸

TLCAF uses the faster R-CNN network for feature extraction and uses faster R-CNN's detection model to

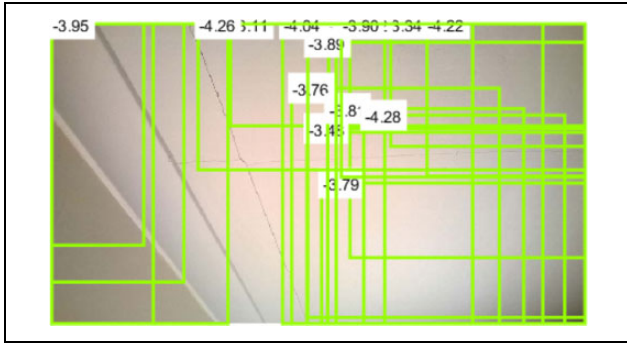


Figure 3. Top-N ranking ROI proposals (before active TLCAF learning). ROI: region of interest; TLCAF: transfer learning on convolutional activation feature.

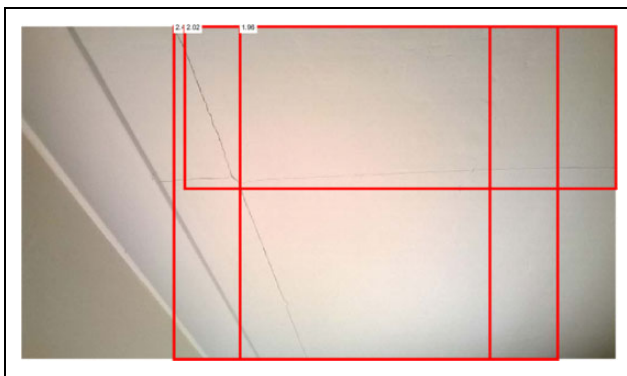


Figure 4. Active TLCAF learning result. TLCAF: transfer learning on convolutional activation feature.

generate the data set for classification. The performance of transferred features is evaluated using linear SVM.

The training and testing steps for the proposed TL network on faster R-CNN are as follows:

Training steps follow faster R- CNN → extraction of features (4096 byte from fully connected layer (fc) 7) → linear classifier (e.g. SVM) → the trained model.

Testing steps are faster R- CNN → extraction of features with regions presented by bbox in fc7 → linear classifier (SVM) with pretrained model for prediction → detection by established regression threshold.

One more fine-tuning process is required for active learning of top-N ranking ROIs to improve recognition rate. A larger “ N ” helps locate visual defect by more precise bboxes. Usually, “ $N = 3$ ” is sufficient for the machine to learn the defect features. In this study, a large value of “ N ,” that is, “ $N = 20$,” was employed to filter out the ambient texture content combination and extract useful features, such that the just right visual defect regions can be identified. Figure 3 presents an example of the detection results before the active learning, and Figure 4 shows the results after the active learning. Comparison of the two

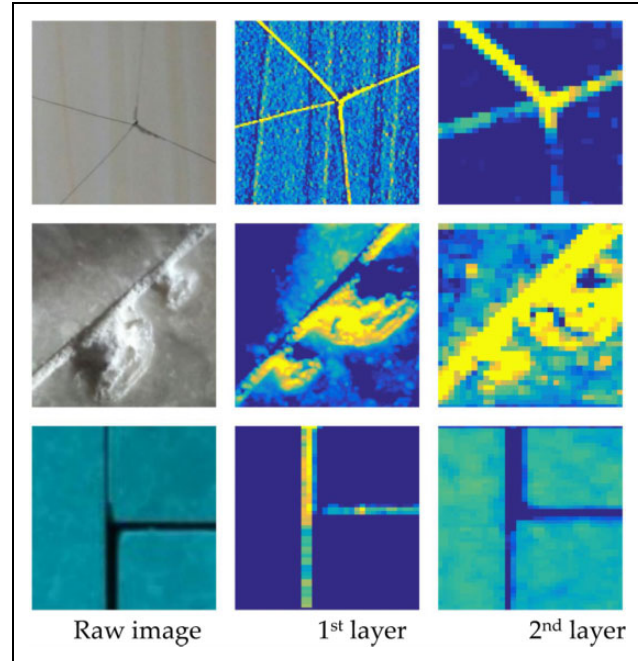


Figure 5. Feature map of damages and jointing defects.

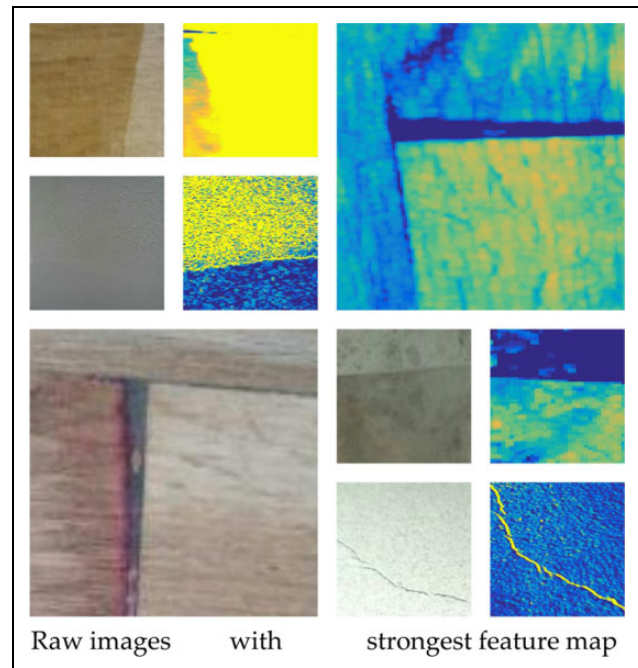


Figure 6. Feature map of finishing defects.

figures indicates significant improvement in accuracy for defect detection achieved by the active learning.

Feature visualization

In Zeiler and Fergus,³⁷ the visualization of features in different convolutional layers indicates that the first two layers’ feature maps project low-level edges and simple

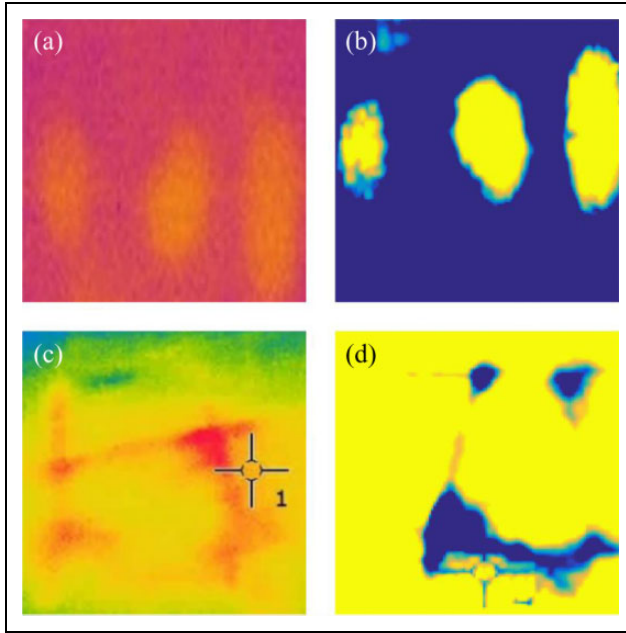


Figure 7. Hollowness feature map.

patterns, and the feature maps of layers three and above present high-level pattern combination. The crack and finishing defects result in sudden gradient decent on edge, color, and/or texture, most of these features are low level ones. Figure 5 shows the typical feature map of the first and second convolutional layers for damages and jointing defects. It demonstrates that the second convolutional layer can filter out more noise and extract useful features. Figure 6 presents some typical defects with their strongest feature maps. In the experiment, horizontally flipping images is used for data augmentation. It shows that the filters inside faster R-CNN can extract the defect feature information with high SNR.

Experimental result

Hollowness detection

A preliminary study to assess the performance of thermal imaging cameras in hollowness detection is conducted using an FLIR A310 infrared camera. Figure 7(a) shows hollownesses detected under ceramic tiles for the indoor thermal images. The ceramic tiles (without reflection) are warmed up for 20 s using a 3 kW Quartz heater via periodic excitation's lock-in IRT method. As the heat conductivity of the ceramic material is high, long-pulse exposure induce nonuniform heating of ceramic tiles and lead to misinterpretation of the results. Hence, sliding window method is employed to minimize the influence of nonuniform heating. In the thermogram, lighter cells indicate there is hollowness beneath the cell, because these cells are less dense with less thermal mass, and thus warmed up faster compared to the solid filled and denser cells. The thermal image

obtained in the experiment demonstrates that thermal response is effective for hollowness detection. Figure 7(c) shows hollownesses and jointing defect detected under granite floor in CONQUAS room after continuously heating up the tiles for 5 min. Figure 7(b) and (d) presents the results of stronger feature map in the second layer of faster R-CNN for classification.

To evaluate the limitation of thermal approach for lighting and reflection, it is tested for ceramic wall with reflection. Similar to ceramic tiles without reflection, hollowness could be identified in the thermal image without heating up, but clearer and more convincing results on hollowness size and shape are available via heating up of the tiles. The experiment results for reflected ceramic wall and granite floor are presented in Figure 8. Figure 8(a) shows the thermal image captured at room temperature (without heating up)—the middle size round hollowness and the big long-shape hollowness (diameter ≥ 8 cm) can be observed, but the third small hollowness, whose diameter is ≤ 5 cm and considered as nonsignificant, is not obviously visible. Figure 8(b) shows the results after heating up for 3 min—a middle size round hollowness and a big oval hollowness can be clearly observed with correct dimensions, the third small hollowness can be observed, but it is not as clear as the bigger ones. The third small hollowness is considered as nonsignificant because of its small size. It can be seen from these figures that the results with proposed methods are quite promising.

In the second experiment, granite floor built in test bed with hollowness under it is tested. A 3 kW Quartz heater and LPT are utilized to continuously heat the tiles for 5 min, and the hollowness and bounding defects hidden beneath the surface are captured in the thermal image, as shown in Figure 8(c). Figure 8(d) displays the hollowness defects with different sizes detected under granite floor after heating up for 5 min. These results indicate that IRT is a reliable tool for hollowness detection.

Classification and detection of defects

Color images are used for detection and assessment of crack and finishing defects as these defects cannot be observed through thermography.

Using the approach proposed in “Methodology” section, the building visual defects are visualized in Figure 9 using Principal Component Analysis (PCA). PCA performs dimensionality reduction by projecting the whole training on a subspace which maximizes the information of data. However, the projection is not the best for classification in this study; it mixes common features with defect features, which results in a significant drop in accuracy. Therefore, PCA is only employed to visualize the top three principal components (PCs) of 2000 random observations, with objective to observe the separability of classes, as shown in Figure 9. In Figure 9, positive features are shown as yellow dots in the 3-D space of the top three PCs, and blue dots represent

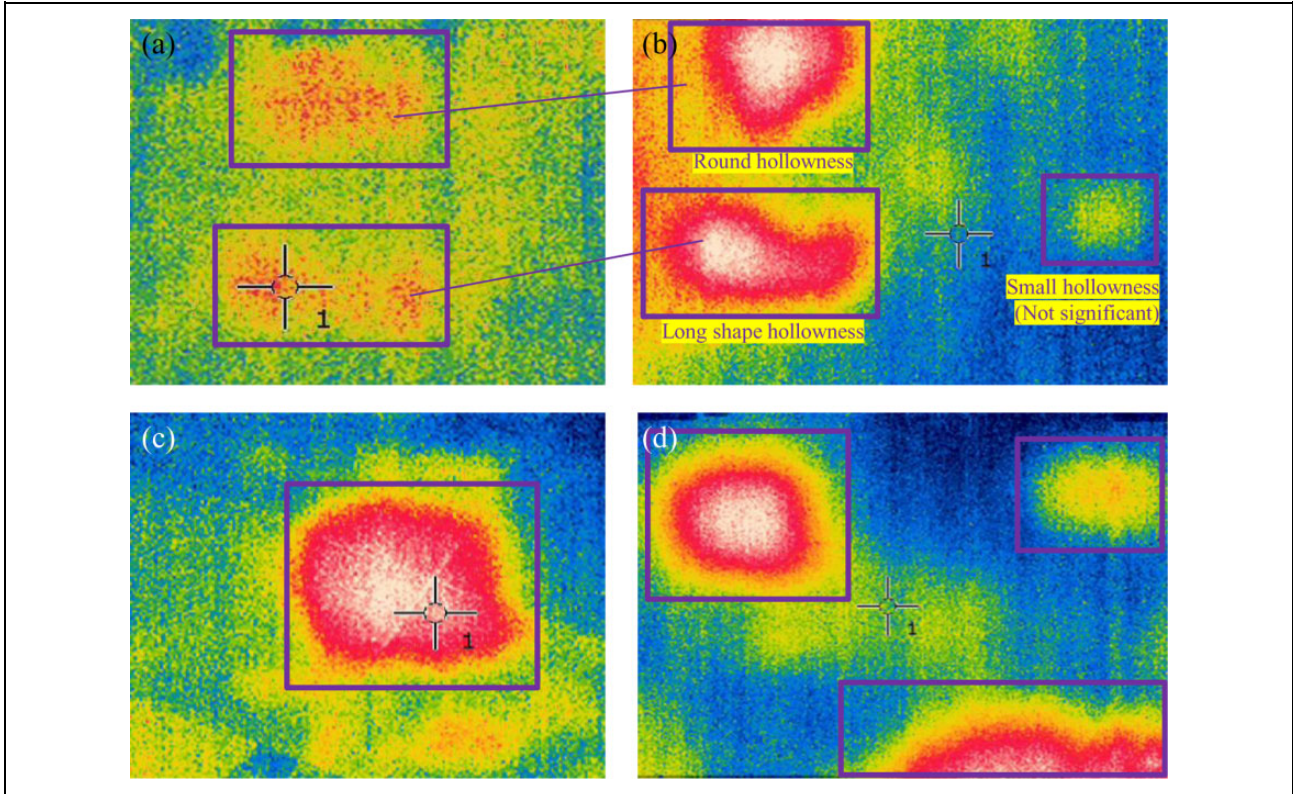


Figure 8. Thermal images of hollowness. (a) Ceramic tiles with reflection (room temperature), (b) ceramic tiles with reflection (heat up 3 min), (c) granite floor with hollowness and bounding defects (heat up 5 min), and (d) granite floor—with three hollowness (heat up 5 min).

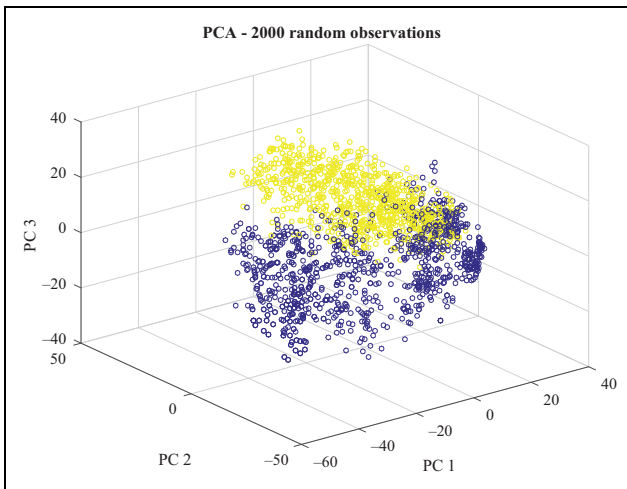


Figure 9. Visualization of visual defect features.

the negative features. Histogram of PC 3 represents good separability of classes between positive features and non-positive features.

In this data set, 1041 features are extracted from 680 images with crack and finishing defects. These features are used as positive set for visual defects of a building sector, as crack and finishing defects share some common low-level features, including sudden gradient change and

texture unevenness. Among the 1041 features, 80% of them are randomly put into the training set, and the rest 20%, that is, 201 features, are put into the testing set. For nonpositive features, 49,781 features are extracted from 610 indoor images without building sector visual defects and employed for the purpose; 80% of the features are randomly put into the training set, and the rest 20%, that is, 9922 features, are put into the testing set. Twenty-five runs are tested. In the experiment, balanced F1 measure (F_{score}) is employed for representing the harmonic mean of precision and recall (it ranges from 0% to 100%), higher F_{score} means better performance. The best model achieved 99.99% accuracy rate, with F_{score} of 99.76%. The mean accuracy rate of the 25 randomly trained TLCAF models is $99.62\% \pm 0.33\%$, and the mean F_{score} is $91.64\% \pm 6.49\%$.

The success measure is defined below where tp is true positive, fp stands for false positive, fn represents false negative, and tn shows true negative

$$\text{Accuracy} = \frac{tp + tn}{tp + fp + fn + tn} \quad (1)$$

“Accuracy” is the most popular measure in machine learning field, especially for deep learning. Nevertheless, it does not discriminate between the numbers of correct labels for different classes. On the contrary, sensitivity and

Table 1. F-measure of building sector's visual defects classification.

TLCAF	Accuracy (%)	Sensitivity (%)	Specificity (%)	Precision (%)	F_score (%)
The best model	99.99	99.51	100	100	99.76
Mean of the 25 randomly Trained TLCAF models	99.62 ± 0.33	96.79 ± 2.55	99.68 ± 0.35	87.76 ± 11.13	91.64 ± 6.49

TLCAF: transfer learning on convolutional activation feature.

Table 2. Compare of the classifier's abilities by the new measure.

Classifier for crack detection		Sensitivity (%)	Specificity (%)	Avoidance of failure		Confirmation of classes for positives		Confirmation of classes for negatives		DP	
				γ	Level	ρ^+	Level	ρ^-	Level	DP	Level
TLCAF (with finishing defects)	Best	99.51	100	0.9951	Superior	+Infinity	Superior	0.0049	Superior	+Infinity	Good
	Mean	96.79	99.68	0.9647	Superior	302.47	Superior	0.0322	Superior	2.19	Fair
Wilk's λ analysis + pyramid SVM ¹		96	98	0.94	Middle-level	48	Middle-level	0.0408	Middle-level	1.69	Limited
Adaptive crack detection ²	LDA + NN	84.1	74.5	0.586	Inferior	3.3	Inferior	0.213	Inferior	0.66	Poor
	LDA + SVM	84.1	72	0.561	Inferior	3	Inferior	0.2208	Inferior	0.63	Poor

DP: discriminant power; TLCAF: transfer learning on convolutional activation feature; SVM: support vector machine; LDA: linear discriminant analysis; NN: neural network.

specificity provide measures that separately evaluate the performance of a classifier for different classes. Sensitivity represents the true positive ratio and specificity shows the true negative proportion³⁹

$$\text{Sensitivity} = \frac{tp}{tp + fn} \quad (2)$$

$$\text{Specificity} = \frac{tn}{fp + tn} \quad (3)$$

In the balanced F1 measure, precision is defined as the percentage of true positives to all predicted positives, and sensitivity is defined as the rate of true positives to all actual positives

$$\text{Precision} = \frac{tp}{tp + fp}; \text{ recall} = \text{sensitivity} = \frac{tp}{tp + fn} \quad (4)$$

The F1 score is expressed by

$$F_{\text{score}} = \frac{2 \times \text{precision} \times \text{recall}}{(\text{precision} + \text{recall})} = \frac{2 \times tp}{2 \times tp + fp + fn} \quad (5)$$

The values of F_{score} listed in Table 1 demonstrate that the proposed network yields high F_{score} . Moreover, an additional performance measure is employed, that is, the DP.

Youden's index γ is employed to evaluate the capability of the algorithm to avoid failure, and it equally weights the performance on positive and negative observations³⁹

$$\gamma = \text{sensitivity} - (1 - \text{specificity}) \quad (6)$$

A higher positive likelihood ρ^+ means better performance on positive class, and a lower negative likelihood ρ^- indicates better performance on negative class

$$\rho^+ = \text{sensitivity}/(1 - \text{specificity}) \quad (7)$$

$$\rho^- = (1 - \text{sensitivity})/\text{specificity} \quad (8)$$

DP evaluates the performance of an algorithm in discriminating between negative and positive samples, and it is usually used in machine learning for feature selection. For an algorithm, it is a poor discriminant if $DP < 1$, a limited one if $DP < 2$, a fair one for $DP < 3$, and a good one in other cases³⁹

$$DP = \frac{\sqrt{3}}{\pi} \times (\log X + \log Y) \quad (9)$$

$$X = \frac{\text{sensitivity}}{1 - \text{sensitivity}}; Y = \frac{\text{specificity}}{1 - \text{specificity}} \quad (10)$$

In the study by Jahanshahi and Masri,² the crack segmentation parameters are adjusted in an adaptive manner based on depth parameters, and the crack features are extracted using LDA. The performance of an SVM, a nearest-neighbor classifier, and an NN is evaluated for classifying cracks from noncrack patterns. This led to the development of an autonomous crack quantification approach using the obtained crack map, which can be applied for detection of crack thickness. In the study by Nashat et al.,¹ pyramid SVM after Wilk's λ analysis is used for crack inspection of biscuits, and the



Figure 10. Detection result.

effectiveness of Hough-based features for crack detection is demonstrated.

The capabilities of different approaches for crack classification are compared in Table 2. The proposed TLCAF network (RPN + SVM) yields the highest values for γ , indicating that it is the best in avoiding failure when compared to other crack detection approaches. The higher $\rho+$ and lower $\rho-$ of the TLCAF network prove that it provides superior performance for confirmation of both positive and negative observations. The DP values indicate the good performance of the TLCAF model, with the random trained TLCAF model a fair discriminant, and the best TLCAF model a good discriminant. In contrast, Wilk's γ analysis + pyramid SVM method yields limited DP, and adaptive crack detection methods using LDA + NN/SVM only provide poor DP.

Figure 10 illustrates the defect detection results after integrating RPN and TLCAF; 53 positive images and 166 nonpositive images are tested, and it yields 93.15% accuracy and 98.09% specificity. The balanced 80.65% sensitivity is acceptable to provide more false positive images than lose important true positive images. Note that these are the results before fine-tuning. After fine-tuning, obvious improvement in accuracy is achieved—there are 15 images that cannot be recognized by TLCAF; all these images are well recognized and detected after applying the A-TLCAF active learning process.

Discussion and future work

The active TL network proposed in this study for defect detection uses RPN for ROI proposal, deep learning network for feature extraction, and linear classifier with supervised learning as object prediction. Moreover, “active transfer learning of top-N ranking regions of interest” is incorporated for fine-tuning of the proposed TL network, and it aims to improve the recognition rate of

visual defects. It is discovered that the pretrained convolutional feature maps produce excellent results, and the use of fine-tuning further increases the accuracy rate of defect detection. The good agreement between experimental results and theoretical values demonstrates that the A-TLCAF network can help improve the system's intelligence level.

Autonomous postconstruction quality assessment is a promising research topic. Current visual assessment of crack and finishing defects for building sectors by human inspectors is highly qualitative and subjective, and the sampling method might miss some area with defects. This leads to imperative need for developing autonomous approaches for building sector inspection. In this study, a novel TLCAF damage/defect detection approach is developed. It is based on an active TL network where deep learning network is used for detection and verification of visual defects, including hollowness, crack, and finishing defects. The capabilities and limitations of the approach are elaborated throughout experiments. Compared with manual assessment, the proposed approach yields higher accuracy and better efficiency for building inspection, and it is also appropriate for integration with fully autonomous and/or semiautonomous mobile robot systems. Moreover, the proposed network can also be employed to extract and train the features of hollowness defects. This requires tests on more samples to build data set and verify during on-site test.

As a part of future work, BIM will be integrated into the proposed network to show the exact defect location based on the accurate localization and mapping information provided by the mobile robot and the sensors. To the best of the authors' knowledge, this study is the first attempt toward developing an autonomous inspection system for defect detection for building sectors. The findings of this investigation provide new insights into this area and lead to an encouraging future research.

Acknowledgements

The authors would like to thank Dr Wang Anran, Mr William Gu, and Mr Tan Wei Chuan for their insightful comments and suggestions. Thanks to Ms Jayanthi Peariahamsy, Mr Ng Kian Wee, Mr Low Chin Leong, Mr Lee Jin Long, and Mr Bryan Soh for their support and advice. Moreover, the support of BCA Academy for sharing domain knowledge is greatly appreciated.

Declaration of conflicting interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This work is funded by National Research Foundation (NRF2015-TDIR01-03), supported by Jurong Town Corporation, collaborated with CtrlWorks, and performed at Nanyang Technological University.

References

- Nashat S, Abdullah A and Abdullah MZ. Machine vision for crack inspection of biscuits featuring pyramid detection scheme. *J Food Eng* 2014; 120: 233–247.
- Jahanshahi MR and Masri SF. Adaptive vision-based crack detection using 3D scene reconstruction for condition assessment of structures. *Automat Constr* 2012; 22: 567–576.
- Maierhofer C, Arndt R, Rollig M, et al. Application of impulse-thermography for non-destructive assessment of concrete structures. *Cement Concrete Comp* 2006; 28(4): 393–401.
- Hung YY, Chen YS, Ng SP, et al. Review and comparison of shearography and active thermography for nondestructive evaluation. *Mat Sci Eng R* 2009; 64(5–6): 73–112.
- Le M, Lee J, Jun J, et al. Hall sensor array based validation of estimation of crack size in metals using magnetic dipole models. *NDT and E Int* 2013; 53: 18–25.
- Brown JR and Hamilton HR. Quantitative infrared thermography inspection for FRP applied to concrete using single pixel analysis. *Constr Build Mater* 2013; 38: 1292–1302.
- Prasanna P, Dana KJ, Gucunski N, et al. Automated crack detection on concrete bridges. *IEEE Transactions on Automation Science and Engineering*. 2016; 13(2): 591–599.
- AL-Marakeby A, Aly AA and Salem FA. Fast quality inspection of food products using computer vision. *Adv Res Comp Commun Eng* 2013; 2(11): 4168–4171.
- Bu G, Chanda S, Guan H, et al. Crack detection using a texture analysis-based technique for visual bridge inspection. *Electr J Struct Eng* 2015; 14(1):41–48.
- Chen Z, Derakhshani R, Halmen C, et al. A texture-based method for classifying cracked concrete surfaces from digital images using neural networks. In: *Neural networks (IJCNN), The 2011 international joint conference*, San Jose, CA, USA, 31 July 2011, pp. 2632–2637. IEEE.
- Ahmed R, El Sayed M, Gadsden SA, et al. Automotive internal-combustion-engine fault detection and classification using artificial neural network techniques. *IEEE Trans Vehicular Technol* 2015; 64(1): 21–33.
- Shafi'i MA and Hamzah N. Internal fault classification using artificial neural network. In: *Proceedings of 2010 4th international power engineering and optimization conference*. Shah Alam, Selangor, Malaysia, 23 June 2010, pp. 352–357.
- Li B, Zhu X, Zhao S, et al. HV power equipment diagnosis based on infrared imaging analyzing. In: *Proceedings of international conference on power system technology*, Chongqing, China, 22 October 2006, pp. 1–4.
- Rahmani A, Haddadnia J and Seryasat O. Intelligent fault detection of electrical equipment in ground substations using thermo vision technique. In: *Proceedings of 2010 2nd international conference on mechanical and electronics engineering*, vol. 2, Kyoto, Japan, 1 August 2010, pp. 150–154.
- Almeida CAL, Braga AP, Nascimento S, et al. Intelligent thermographic diagnostic applied to surge arresters: a new approach. *IEEE Trans Power Deliv* 2009; 24(2): 751–757.
- Chou YC and Yao L. Automatic diagnostic system of electrical equipment using infrared thermography. In: *International conference of soft computing and pattern recognition*, Malacca, Malaysia, 4 December 2009, pp. 155–160.
- de Oliveira JHE and Lages WF. Robotized inspection of power lines with infrared vision. In: *Applied robotics for the power industry (CARPI), 2010 1st international conference*, Montreal, QC, Canada, 5 October 2010, pp. 1–6.
- Lim RS, La HM, Shan Z, et al. Developing a crack inspection robot for bridge maintenance. In: *2011 IEEE international conference on robotics and automation (ICRA)*, Shanghai, China, 9 May 2011, pp. 6288–6293.
- Sharifi M, Fathy M and Mahmoudi MT. A classified and comparative study of edge detection algorithms. In: *Proceedings of international conference on information technology: coding and computing*, Las Vegas, NV, USA, 8 April 2002, pp. 117–120. IEEE.
- Girshick R, Donahue J, Darrell T, et al. Rich feature hierarchies for accurate object detection and semantic segmentation. In: *Proceedings of the IEEE conference on computer vision and pattern recognition*, 24 June 2014, pp. 580–587.
- Uijlings JRR, van de Sande KEA, Gevers T, et al. Selective search for object recognition. *Int J Comput Vis* 2013; 104(2): 154–171.
- Girshick R. Fast R-CNN. In: *Proceedings of the IEEE international conference on computer vision*, 13 December 2015, pp. 1440–1448.
- Alexe B, Deselaers T and Ferrari V. Measuring the objectness of image windows. *IEEE Trans Pattern Anal Mach Intell* 2012; 34(11): 2189–2202.
- Zitnick CL and Dollár P. Edge boxes: locating object proposals from edges. In: Fleet D, Pajdla T, Schiele B, et al. (eds) *Computer vision – ECCV 2014. Lecture Notes in Computer Science*. Cham: Springer, 2014, vol. 8693, pp. 391–405.
- Arbeláez P, Pont-Tuset J, Barron J, et al. Multiscale combinatorial grouping. In: *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 328–335, 24 June 2014.

26. Carreira J and Sminchisescu C. CPMC: automatic object segmentation using constrained parametric min-cuts. *IEEE Trans Pattern Anal Mach Intell* 2012; 34(7): 1312–1328.
27. Wang A, Lu J, Cai J, et al. Large-margin multi-modal deep learning for RGB-D object recognition. *IEEE Trans Multi-media* 2015; 17(11): 1887–1898.
28. He K, Zhang X, Ren S, et al. Spatial pyramid pooling in deep convolutional networks for visual recognition. *IEEE Trans Pattern Anal Mach Intell* 2015; 37(9): 1904–1916.
29. Hosang J, Benenson R, Dollár P, et al. What makes for effective detection proposals? *IEEE Transactions on Pattern Analysis and Machine Intelligence* 2016; 38(4): 814–830.
30. Ren S, He K, Girshick R, et al. Faster R-CNN: towards real-time object detection with region proposal networks. In: Cortes C, Lawrence ND, Lee DD, et al. (eds) *Advances in neural information processing systems*. 2015, pp. 91–99.
31. Liu W, Anguelov D, Erhan D, et al. SSD: single shot multi-box detector. In: *European conference on computer vision*, pp. 21–37. Springer, 2016.
32. Torrey L and Shavlik J. Transfer learning. In: Olivas ES, Guerrero JDM, Sober MM, et al. (eds) *Handbook of research on machine learning applications and trends: algorithms, methods, and techniques*. vol. 1, 2009, p. 242.
33. Pan SJ and Yang Q. A survey on transfer learning. *IEEE Trans Knowl Data Eng* 2010; 22(10): 1345–1359.
34. Tai L, Ye Q and Liu M. PCA-aided fully convolutional networks for semantic segmentation of multi-channel fMRI. *arXiv preprint arXiv:1610.01732*, 2016.
35. Chorowski JK, Bahdanau D, Serdyuk D, et al. Attention-based models for speech recognition. In: Cortes C, Lawrence ND, Lee DD, et al. (eds) *Advances in neural information processing systems*. 2015, pp. 577–585.
36. Szegedy C, Reed S, Erhan D, et al. Scalable, high-quality object detection. *arXiv preprint arXiv:1412.1441*, 2014.
37. Zeiler MD and Fergus R. Visualizing and understanding convolutional networks. In: Fleet D, Pajdla T, Schiele B, et al. (eds) *Computer vision—ECCV 2014*. Springer, 2014, pp. 818–833.
38. Ren S, He K, Girshick R, et al. Object detection networks on convolutional feature maps. *arXiv preprint arXiv:1504.06066*, 2015, 2015.
39. Sokolova M, Japkowicz N and Szpakowicz S. Beyond accuracy, f-score and ROC: a family of discriminant measures for performance evaluation. In: Sattar A and Kang B (eds) *AI 2006: advances in artificial intelligence*. Berlin, Heidelberg: Springer. 2006, pp. 1015–1021.