Online deep transfer learning applied to building quality assessment robots

Liu, Lili

2019

Liu, L. (2019). Online deep transfer learning applied to building quality assessment robots.
Doctoral thesis, Nanyang Technological University, Singapore.

<https://hdl.handle.net/10356/104468>

<https://doi.org/10.32657/10356/104468>

**NANYANG
TECHNOLOGICAL
UNIVERSITY**

SINGAPORE

**ONLINE DEEP TRANSFER LEARNING
APPLIED TO BUILDING QUALITY ASSESSMENT ROBOTS**

LIU LILI

School of Mechanical and Aerospace Engineering

2019

ONLINE DEEP TRANSFER LEARNING APPLIED TO BUILDING QUALITY ASSESSMENT ROBOTS

LIU LILI

School of Mechanical & and Aerospace Engineering

A thesis submitted to the Nanyang Technological University
in partial fulfillment of the requirements for the degree of
Doctor of Philosophy

2019

Statement of Originality

I hereby certify that the intellectual content of this thesis is the product of my original research work and has not been submitted for a higher degree to any other University or Institution.

Jan. 2019

.....

Date

.....

LIU LILI

Supervisor Declaration Statement

I have reviewed the content and presentation style of this thesis and declare it is free of plagiarism and of sufficient grammatical clarity to be examined. To the best of my knowledge, the research and writing are those of the candidate except as acknowledged in the Author Attribution Statement. I confirm that the investigations were conducted in accord with the ethics policies and integrity standards of Nanyang Technological University and that the research data are presented honestly and without prejudice.

Jan. 2019

.....

Date

.....

Prof. Chen I-Ming

Authorship Attribution Statement

This thesis contains material from three papers published in the following peer-reviewed journals / from papers accepted at conferences in which I am listed as an author.

Chapter 3 is published as [L. Liu, R.-J. Yan, V. Maruvanchery, E. Kayacan, I.-M. Chen, L. K. Tiong, Transfer learning on convolutional activation feature as applied to a building quality assessment robot, International Journal of Advanced Robotic Systems 14 \(3\) \(2017\) 1729881417712620.](#)

The contributions of the co-authors are as follows:

- Prof. Chen I-Ming and Prof. Erdal Kayacan provided the initial project direction and edited the manuscript drafts.
- I prepared the manuscript drafts. The manuscript was revised by Prof. Erdal Kayacan, Prof. Tiong Lee Kong, Robert, Dr. Yan Rui-Jun and Mr. Varun Maruvanchery.
- I co-designed the active thermal graphic study with Prof. Erdal Kayacan and Mr. Varun Maruvanchery and performed all the laboratory work at the School of Mechanical and Aerospace Engineering and School of Civil and Environmental Engineering. I analyzed the data.
- I designed a new image-based real-time post-construction quality assessment system to increase the effectiveness and reliability of the assessment. Active transfer learning for convolutional activation feature (A-TLCAF) network design and experiments, including data preparation, were conducted by me for image-based cracks, finishing defects and hollowness assessment in the School of Mechanical and Aerospace Engineering, data are collected online, onsite and from testbeds.
- Dr. Yan Rui-Jun supports on mechanical design and implementation on A-CONQUAS robot system integration, alignment and evenness assessment.
- Dr. Varun Maruvanchery assisted in the testbed preparation according to our project requirement.

Chapter 4.2 is published as [L. Liu, I.-M. Chen, E. Kayacan, L. K. Tiong, V. Maruvanchery, Automated construction quality assessment: A review, in: Mechatronics and its Applications \(ISMA\), 2015 10th International Symposium on, IEEE, 2015, pp. 1-6. DOI: 10.1109/ISMA.2015.7373459.](#)

The contributions of the co-authors are as follows:

- Prof. Chen I-Ming provided the initial project direction and edited the manuscript drafts.

- I prepared the manuscript drafts. The manuscript was revised by Prof. Erdal Kayacan, Prof. Tiong Lee Kong, Robert and Mr. Varun Maruvanchery.
- I studied hollowness detect methods of different materials by active thermography analysis.
- We proposed a robotic platform, which integrates control, sensing and drive, to achieve an intelligent and automated quality assessment system.

Chapter 4.4 is published as [L. LIU, E. Tan, Z. Q. Cai, X. J. Yin, Y. Zhen, CNN-based automatic coating inspection system, Advances in Science, Technology and Engineering Systems Journal 3 \(6\) \(2018\) 469-478. DOI: 10.25046/aj030655.](#)

The contributions of the co-authors are as follows:

- Dr. Zhen Yongda and Dr. Cai Zhi Qiang provided the initial project direction and edited the manuscript drafts.
- I wrote the drafts of the manuscript. The manuscript was revised together with Dr. Yin Xi Jiang and Ms. Estee Tan.
- I developed the CNN-based automatic coating inspection system and conducted data evaluation.
- Ms. Estee Tan prepared the testing samples and supported on active thermal graphic experiments.
- Dr. Cai Zhi Qiang designed coating corrosion color dataset, conducted the testing for our coating inspection system and give advices.

Jan. 2019

.....

Date

.....

LIU LILI

Acknowledgements

First of all, from the bottom of my heart, I would like to express my heartfelt thanks to Professor Chen I-Ming, from Nanyang Technological University (NTU), for his patient guidance and continuous encouragement during my Ph.D. study. It is my greatest honor to have him as my supervisor, who led me step by step in becoming a qualified researcher. I want to thank Prof. Erdal Kaycan, Prof. Yeo Song Huat, Prof. Ang Wei Tech and Prof. Seet Gim Lee, Gerald, Prof. Domenico Campolo, Dr. Wang Anran, Prof. Tan Ah Hwee, Prof. Ho Shen-Shyang, Prof. Tiong Lee Kong, Robert and Prof. Lum Guo Zhan for their kind advice and support.

Secondly, I would like to thank Mr. Ng Kian Wee, Dr. Albert Causo, Dr. William Gu, Ms. Emily Toh, Dr. Yan Rui Jun, Ms. Jayanthi Peariahsamy for their insights and comments. Moreover, the support from Jurong Town Corporation (JTC), CtrlWorks and BCA Academy is greatly appreciated. I would like to thank Trimble navigation for their on-site surface measurement experiments. Thanks to Mr. Low Chin Leong for his help in robotic system design, thanks to Dr. Liang Conghui, Dr. Meghdad Attarzadeh, Mr. Muley Pravin Sudhakar, Mr. Law Weichuan and Dr. Tan Weichian for their support.

Thirdly, I would like to thank all the great people I work with in Robotic Research Center, NTU, Singapore. I would like to thank Prof. Chen I-Ming for taking me into this world-class research team to pursue valuable research goals. Many thanks to Prof. Lyu, Chen and three other anonymous examiners for their sincere advice and support. I would like to thank all the classmates, colleagues, friends for their kind help and the fruitful collaboration. Special thanks to Dr. Yuan Qilong and Dr. Guo Wenjiang, Mr. Li Bingbing, Dr. Ang Wei Sin for their advice, assistants, friendship and encouragement. I would especially like to thank Dr. Zhen Yongda, Dr. Cai Zhiqiang, Dr. Yin Xijiang, Dr. Rajnish Gupta, Mr. David Chai, Dr. Tao Naifu, Mr. Edwin Loh and Miss Estee Tan for their kind support from Singapore

Polytechnic, thank Dr. Gu Hai and the team support from American Bureau of Shipping.

Fourthly, I am very grateful to NTU for providing me with valuable research opportunities to pursue my Ph.D. study. The research was funded in part by National Research Foundation (NRF2015-TDIR01-03), A* Star SERC, and SMI (SMI2015-OF-05), collaborated with JTC and CtrlWorks, conducted at NTU.

Finally, I am very grateful to my family for their support. They provided me with a balanced life, education, and constant encouragement. I cherish their love and support.

Abstract

Post-construction quality assessment is critical to the building projects. It is labour intensive and time consuming. The results of the assessment depend on the examiner performing the assessment and are therefore subjective people may have various opinions about an assessment and people may make mistakes; therefore, different examiners may give different results. Recent development in artificial intelligence techniques has enable design of an automated system for building quality assessment, to increase objectivity and accuracy, and reduce labour costs. This motivated current research, thereby establishing an automated post-construction quality assessment system for detecting various types of defects, such as cracks, finishing defects and hollowness. Compared to traditional methods, the system greatly reduces labour costs and provides a fast, objective and accurate assessment. In the proposed system, transfer learning for convolutional activation feature (TLCAF) networks, active-TLCAF (A-TLCAF) and online-TLCAF networks are employed for task automation. In the TLCAF network, faster R-CNN test mode is used as the base model for the proposal of region of interest (ROI), and a deep transfer learning (DTL) network is employed for model training and defect classification; finally, non-maximum suppression (NMS) and threshold adjustment are performed for defect detection. The active TLCAF (A-TLCAF) network allows users to actively intervene the labelling work of the top-N ROIs, and fine-tune the networks using the newly labelled images. Compared with TLCAF, A-TLCAF can improve the detection accuracy.

To improve learning speed and to achieve incremental learning without forgetting existing knowledge, an online deep transfer-learning network is also proposed. The network is termed as online-TLCAF, whereby YOLO is used as the underlying network to deliver generic objects, convolutional neural networks are employed for extraction of features of visual defects, and broad learning algorithm is used for incremental learning. The system provides generalization capabilities for function approximation and simplifies the final structure using singular value decomposition (SVD). Compared with TLCAF, online-TLCAF has two improvements: 1).

the ROI proposal network is replaced by an automated object proposal, which eliminates the need for ROI labelling work; 2). the linear classifier in TLCAF is replaced by an online learning system. The online-TLCAF network proposed in this study provides an incremental learning for high-dimensional dynamic image/video streams.

Extensive experiments were conducted in the CONQUAS room, the test bed and our self-built data set, and the results were used to validate the developed automated post-construction quality assessment system. Various learning algorithms have been developed to illustrate the power of the proposed framework. The new method is satisfactory in evaluating various image-based defects. Compared to shallow structures, online-TLCAF provides greater flexibility for image/video-based object detection. Compared to traditional manual inspections, this automated system is suitable for large area inspections and increasing efficiency and reliability.

Contents

Acknowledgements	ix
Abstract	xi
List of Figures	xv
List of Tables	xvii
Acronyms	xix
1 Introduction	1
1.1 Background	1
1.2 Problem Statement	1
1.3 Issues and Challenges	3
1.4 Methodology	4
1.5 Main Contributions	5
2 Literature Review	7
2.1 An Overview on Artificial Neural Networks	7
2.2 Deep Learning - A Deeper Dive into Thinking Mechanism	9
2.2.1 Region Proposal	12
2.2.2 Object Classification	15
2.2.3 Model Optimization and Fine-tuning	17
2.2.4 Object Detection	18
2.3 Transfer Learning	21
2.4 Online Learning	28
2.5 Summary	34
3 Online Deep Transfer Learning	37
3.1 TLCAF	41
3.2 A-TLCAF	41
3.3 Online-TLCAF	45
3.4 Model Evaluation	49
3.4.1 CIFAR-10	49

3.4.2	Building Defect Dataset	54
3.5	Summary	63
4	Technical Contribution for Defect Detection	65
4.1	Crack Detection - Work in NTU	66
4.1.1	Approach	66
4.1.2	Feature Visualization	68
4.1.3	Defect Prediction and Detection	69
4.1.4	Experimental Results	70
4.2	Hollowness Assessment - Work in NTU	72
4.2.1	Approach	74
4.2.2	Experimental Results	76
4.3	A-CONQUARS Robot System	78
4.3.1	Surface Measurement	78
4.3.2	Sensor Fusion and Integration	80
4.3.3	Integrated Robot System	82
4.4	Coating Condition Assessment - Work in Singapore Polytechnic . .	84
4.5	Summary	86
5	Conclusions and Future Work	89
5.1	Conclusions	89
5.2	Future Work	91
	List of Author's Patents and Publications	93
	List of Author's Patents and Publications	93
	Bibliography	97

List of Figures

1.1	A-CONQUARS system (A-CONQUARS)	2
1.2	Quicabot	3
1.3	Issues and challenges	4
2.1	Biological neural networks [1]	8
2.2	An overview of salient object detection [2]	13
2.3	Image labeler to set ground-truth	14
2.4	ImageNet classification challenge	16
2.5	Bias variance tradeoff	17
2.6	Bounding box object detectors: understanding YOLO [3, 4]	20
2.7	Driving force behind the success of machine learning industry[5]	23
2.8	Traditional supervised learning settings in ML	24
2.9	Transfer learning settings	25
2.10	TL classification (based on content transferred)	27
2.11	Overview of the different settings of transfer learning [6]	28
2.12	Illustration of inductive transfer	28
3.1	Online-TLCAF	38
3.2	Categorization of DTL [7]	39
3.3	DTL fine tuning procedures	41
3.4	TLCAF	42
3.5	A-TLCAF	42
3.6	A-TLCAF network	44
3.7	Online-TLCAF for image-based defect detection	46
3.8	YOLO objectness proposal [3, 4]	47
3.9	Objectness proposals for image-based building defect	48
3.10	VGG19-based TLCAF (CIFAR-10 dataset)	50
3.11	Layer fc features (CIFAR-10 dataset)	50
3.12	ResNet-50 based TLCAF (CIFAR-10 dataset)	51
3.13	Labeled building defects: crack, corrosion, non-defect and finishing defect	54
3.14	Deep transfer learning effect (our building defect dataset)	56
3.15	Comparison among different network-based TLCAF for defect validation accuracy	56
3.16	Visualization of defect feature	58

3.17	ResNet-50 based TLCAF for building defects recognition	58
3.18	ROC of one vs others	59
3.19	Comparison of complexity for different CNN models [8]	60
3.20	CIFAR-10 selected dataset	61
3.21	Visualization of TLCAF trained CIFAR-10 features through T-SNE	61
4.1	A-TLCAF learning	68
4.2	Feature map of different defects	69
4.3	Visualization of image-based defect features	70
4.4	Detection result [9]	70
4.5	Hollowness feature map	76
4.6	Thermal images of hollowness	77
4.7	Hollowness detected	78
4.8	Perpendicularity measurement	79
4.9	Flatness measurement	79
4.10	Quicabot v1 and selected sensors [10]	83
4.11	Assessment results of QuicaBot for image-based defects	84
4.12	Instance-aware semantic segmentation [11]	85
4.13	Feature prediction results [11]	86

List of Tables

2.1	Differences between offline and online learning	30
2.2	Comparison of online learning algorithms	32
3.1	ResNet-50 based Online-TLCAF results	53
3.2	Validation accuracy vs. training time for different ResNet architectures on CIFAR-10 [12]	53
3.3	Validation accuracy vs. training time for different TLCAF models on CIFAR-10	54
3.4	Comparison between online and offline models	60
3.5	Different modes for online-TLCAF	62
3.6	Pre-trained VGG19-based online-TLCAF	62
3.7	Trained VGG19-based online-TLCAF	62
3.8	Pre-trained Resnet-50-based online-TLCAF	63
3.9	Trained Resnet-50-based online-TLCAF	63
4.1	Comparison of crack identification algorithms	71
4.2	Crack detection accuracy	71
4.3	Hollowness detection methods review	75

Acronyms

AI	Artificial Intelligence
ANN	Artificial Neural Network
A-TLCAF	Active Transfer Learning for CAF
BBoxes	Bounding Boxes
BCA	Building and Construction Authority
BIM	Building Information Model
BING	Binarized Normed Gradients
BL	Broad Learning
BLS	Broad Learning System
BP	Back Propagation
CAF	Convolutional Activation Feature
CBC	Coating Breakdown and Corrosion
CIFAR	Canadian Institute For Advanced Research
CNN	Convolutional Neural Networks
CONQUA	Automated Construction Quality Assessment
CONQUAS	Automated Construction Quality Assessment System
A-CONQUARS	Automated CONQUA Robotic System
DP	Discrimination Power
D-S Theory	Dempster-Shafer Theory
DTL	Deep Transfer Learning
EKF	Extended KalmanFilter

faster R-CNN	faster Region-based CNNs
FCN	Full Convolutional Network
GAN	Generative Adversarial Network
GPR	Ground-Penetrating Radar
HOG	Histogram of Oriented Gradients
IELM	Incremental Extreme Learning Machine
ILSVRC	ImageNet Large Scale Visual Recognition Competition
ILVQ	Incremental Learning Vector Quantization
IoU	Intersection of Union
IR	Infrared Radiation
IRT	Infrared Thermography
ISL	Invisible Structure Light
ISVM	Incremental Support Vector Machine
KF	Kalman Filter
Laser UT	Laser-Ultrasonic Testing
LDA	Linear Discriminant Analysis
LPT	Long Pulse Exposure IRT
MCG	Multiscale Combinatorial Grouping
MFI	Multi-sensor Fusion and Integration
NB _{Gauss}	Naive Bayes
NDT	Non-Destructive Test
NIPS	Neural Information Processing Systems
NMS	Non Maximum Suppression
NN	Neuron Networks
online-TLCAF	Online Transfer Learning for CAF
ORF	Online Random Forest
PC	Principal Components
PCA	Principal Component Analysis

PF	Particle Filter
Quicabot	Quality Inspection and Assessment Robot
R-CNN	Region-based CNNs
ReLU	Rectifying Linear Unit
ResNet	Residual Neural Network
RF	Random Forest
RNN	Recurrent Neural Networks
ROI	Region Of Interest
RPN	Region Proposed Network
RVFLNN	Random Vector Functional-Link Neural Networks
SGD	Stochastic Gradient Descent
SGDM	Stochastic Gradient Descent with Momentum
SIFT	Scale-Invariant Feature Transform
SLFN	Single Layer Feed-forward Network
SLIC	Simple Linear Iterative Clustering
SPPnets	Spatial Pyramid Pooling
SVD	Singular Value Decomposition
TL	Transfer learning
TLCAF	Transfer Learning for CAF
t-SNE	t-Distributed Stochastic Neighbor Embedding
UKF	Unscented Kalman Filter
ULT	Ultrasonic Test
VGG	Visual Geometry Group
VOC	Visual Object Classes
YOLO	You Only Look Once
ZFNet	Zeiler and Fergus network

Chapter 1

Introduction

1.1 Background

In the past few years, the pursuit of artificial intelligence (AI) has led to major innovations, especially those in a field called “machine learning”. The main motivation and driving force in these research areas is that the solutions for some problems are too complex/difficult/laborious for manual operation; machine learning techniques are increasingly utilized to solve such problems.

There has been some major progress in visual tasks such as image recognition, object detection, auto-driving car and others that only appear in dreams decades ago. As part of AI, training of deep neural networks from massive labelled data (e.g. images, words, etc), to achieve accurate mappings from input to output is successful. However, it remains a challenge to generalize the model for application in conditions that are different from those encountered in training. The ability to transfer knowledge to new conditions is called transfer learning (TL). It is a machine learning method in which models developed for previous tasks are reused in subsequent tasks.

1.2 Problem Statement

Post-construction quality assessment is crucial for a building. It is labor intensive and time consuming, recent rapid advancements in AI and machine learning

FIGURE 1.1: A-CONQUARS system (A-CONQUARS)

have inspired researchers to explore possibility of automation for the task. Mobile robots, equipped with appropriate sensors, can perform post-construction quality assessment automatically, with a significantly higher assessment speed and a better accuracy, compared to manual inspection. Meantime, the inspection operator is freed from the tedious physical job and able to focus on other tasks. Hence, use of a robotic inspection system generates significant economic benefits, including shorter assessment time, reduced project cost, and also better inspection quality. Moreover, the overall inspection process via the robotic system can be easily recorded for review and assessment later, which is a difficult task for manual inspection operation. The post-construction quality assessment robot system can automatically track defects using predefined algorithms and mark defects on 2D or 3D maps.

The purpose of this work was to develop an online deep transfer learning (DTL) method for automated building quality assessment robotic systems (A-CONQUARS), as shown in Figure 1.1. The quality inspection and assessment robot (QuicaBot), as shown in Figure 1.2, can be used as an assistant to the construction assessor in an automatic manner to speed up construction quality inspections and improve their accuracy. Key requirements of the solution for building defect detection, such as versatility, ease of use, time and cost savings, etc, are addressed in the design and development of the robotic system.

FIGURE 1.2: Quicabot

1.3 Issues and Challenges

Deep learning has been used in many industries and is one of the most advanced AI technologies at the time of this writing. Deep learning algorithms are prone to over-fitting, because of the relative increase in complexity. Meanwhile, the increased model complexity could also lead to a significant/huge demand in computing resources and time. Due to the complexity of combining the model with an optimization algorithm such as gradient descent, it is also important to consider that the solution can represent a local minimum, rather than a global minimum. With all these factors in mind, attention must be taken when solving problems using the AI algorithms, including the choosing of algorithm, implementation and performance evaluation, etc. Deep learning requires a lot of data. In this work, the data for defect detection are limited; therefore, a DTL approach has been explored.

The real-world comes from complex environments that evolve over time. In the online DTL module of this study, a focus is the design of algorithms to handle conceptual drift, i.e. the variation of underlying distribution of data, which causes predictions less accurate over time. To prevent this, a prompt and accurate adaptation of the learning model to changes is required. Therefore, online adaptive machine learning methods are needed to handle the special types of drift of various intensities.

Issues	Goal	Challenges	Advantages
<ul style="list-style-type: none"> • Data are limited • Real-world comes from complex environments that evolve over time 	<ul style="list-style-type: none"> • Aim to think and learn like human • Online DTL is explored 	<ul style="list-style-type: none"> • Optimized initialization • frequent model updates • model maintenance • data contamination 	<ul style="list-style-type: none"> • High-speed data processing capabilities • Able to quickly adapt to data changes

FIGURE 1.3: Issues and challenges

Online learning algorithms face many challenges as shown in Figure 1.3. Many online learning methods require optimized initialization to avoid long convergence times, and frequent model updates can cause unnecessary fluctuations in model output. Furthermore, models trained through online learning methods are often more difficult to maintain, because online models may become unstable or even corrupted, due to data contamination. To maintain the accuracy and stability of the model, a constant monitor of data and model quality is required from the online learning service providers. Despite these challenges, Online learning becomes more valuable because of its high-speed data processing capabilities and the ability to quickly adapt to data changes.

1.4 Methodology

Deep transfer learning model is used to extract common advanced features and transfer them into new tasks. Compared to shallow structures, DTL provides greater flexibility in extracting advanced features and has been proven to promote a variety of scientific and engineering issues [13, 14]. A DTL model was proposed in this study for image-based defect detection, it employs an eight-layer full convolutional network (FCN) for feature extract, a linear classifier for feature learning of different types of defects.

In the proposed online DTL model, analysis is provided specifically for data streams containing concept drift, with an emphasis on algorithms designed to handle special

drift types of various intensities; the transfer learning for convolution activation features (TLCAF) is combined with broad learning algorithms[15] to decrease feature training duration, this provides generalization capabilities for function approximation and simplifies the final structure using singular value decomposition (SVD) simplification.

1.5 Main Contributions

The proposed A-CONQUARS is capable of inspecting cracks, voids, finishing defects, uniformity and alignment for a building. It includes a mobile robot for mapping/navigation, a thermal imaging camera with a heater for hollow detection, a color camera for image-based defect assessment, and a laser scanner and inclinometer for alignment and uniformity inspection. The A-CONQUARS developed can be used to evaluate and better understand defect locations through the Integrated Building Information Model (BIM). This investigation focuses on the detection of cracks, finishing defect and hollowness. The proposed algorithm can be employed to automatically track all types of image-based defects.

The purpose of this study is to design a new image-based real-time post-construction quality assessment system to increase the effectiveness and reliability of the assessment. This method is called online transfer learning for the convolution activation function (online-TLCAF). The novelties of this study are as follows:

- TLCAF and active-TLCAF [9, 11, 16–18] was proposed for image-based defect detection.
- Online transfer learning for convolution activation functions, (online-TLCAF) network was proposed for incremental learning of image-based defect detection.
- Hollownesses below the tiles were studied by active thermography [19–21].
- A robotic platform was designed and developed to achieve an intelligent and automated quality assessment system [19, 22, 23].

The structure of this work is organized as follows. In Chapter 1, a general introduction is given on the background, motivation, challenges, approach and main

contributions of this work. The literature review on online DTL is presented in Chapter 2, and Chapter 3 depicts methodologies, technical details, and mechanisms for online DTL. Subsequently, technical contributions of different DTL scenarios are provided and evaluated in Chapter 4, and a practical approach for online deep transferring learning is developed and used in A-CONQUARS. Finally, Chapter 5 presents an overview of the contributions of this report and future research directions.

Chapter 2

Literature Review

In this chapter, the reasons why transfer learning deserves significant research attention is first described. Following this, the most common and successful example of machine learning, “supervised learning”, is introduced. After that, the advantage of online learning and when to use it is discussed. Subsequently, classification of online learning algorithms is depicted. Finally, the value of online deep transfer learning method is discussed.

2.1 An Overview on Artificial Neural Networks

To define artificial intelligence (AI), the concept of “intelligence” is first explained. Intelligence is primarily about learning, understanding, and applying what is learned to achieve certain goals. In this way, AI is portrayed as the intelligence that the machine displays. Some famous examples of AI solutions include Apple’s Siri, Amazon’s Alexa, Baidu’s xiaodu and the latest Google assistant. AI is also applied in other predictive tasks, including images, text, robots, autonomous vehicles, traffic, crime, medical diagnostics, marketing, search engines, spam filtering, and so on. Among the many different goals of AI, visual intelligence is its cornerstone. An advanced version of the artificial neural network (ANN), deep learning, is the focus of this study.

As part of AI, deep neural networks have witnessed significant innovations through the use of neurophysiologically driven strategies. As a new type of power, AI is

FIGURE 2.1: Biological neural networks [1]

changing every major industry. All industries will be part of this AI-driven future. The work in [24] shows that deep learning can be achieved using isolated dendritic compartments, which may help explain the dendritic morphology of neocortical pyramidal neurons. The models learn to recognize patterns in digital representations of sounds, images, and other data in a real sense. The human brain is a startlingly complex and powerful computing machine. The internal working of human brain is usually modeled by the concepts of neurons and biological neural networks. A human brain contains about 100 billion neurons [25], which are connected along the path of the entire network. At high levels, neurons interact with one other through an interface composed of axonal terminals, which are connected to dendrites that pass through synapse, as shown in Figure 2.1 [1]. If the sum of the weighted input neuron signals exceeds a threshold to cause a message transmission, a single neuron will pass the information to the next neuron through the interface, which is called activation [1]. A single neuron will receive many input signals simultaneously, and these input signals are summed to determine if the message is transmitted along the way, and ultimately the brain is instructed to move, remember memories, and many more. "Thinking mechanism" [26] of the human brain generates instructions to the muscles, organs and body. In addition, the brain's neural network constantly updates itself in different ways, including modification of the weights between neurons. This is a direct impact of learning and experience.

In view of this, it is assumed that to replicate the functions of human brain, including "intelligence", the computer needs to successfully implement a simulated version of the neural network. This is the origin of ANNs. ANNs are statistical models that are directly inspired and partially modeled by biological neural networks. Related algorithms can be used in many applications. The ANN is characterized by including adaptive weights along the path between the neurons, which

can be adjusted by learning algorithms which learn to improve the models from the observed data. Additionally, appropriate cost functions are required for different applications. Cost function is used to learn the optimal solution to a problem. This includes determining the optimal values for all tunable model parameters, which are usually optimized by gradient descent. The capability/accuracy of optimization algorithms determines the performances of ANNs. Structurally, ANNs are modeled using artificial neuron layers that receive input, and a threshold is applied to the activation function to determine if a message is delivered.

The rise in model complexity could lead to over-fitting. In addition to learning algorithms, model architecture and tuning are also important components of artificial neural networks, which have a large impact on model performance. A single neuron represents a local representation. The entire network represents a distributed representation formed by multiple transformations between neurons and layers. Distributed representation is a powerful tool for modeling the semantics of items in a dense, low-dimensional space. It can be used to build advanced machine learning tasks.

2.2 Deep Learning - A Deeper Dive into Thinking Mechanism

AI is about to change our world in ways we can't imagine. AI and computer vision, as the core of machine learning, originate from a landmark experiment of David Hubel and Torsten Wiesel [27] in 1959. Results of the experiment show that brain is connected by neurons, each layer has a hierarchical structure, starting from the simple edges. This groundbreaking work first revealed the computational structure of mammal vision, including human visual system, and they were awarded Nobel Prize in medicine in 1981. Although some psychological theories have already explained the phenomena of cognition and thinking, the biological mechanism and nature of thinking remain a mystery. Zheng and Ma [26] brought forward a new hypothesis for thinking mechanism in 1995. They analyzed sleep, dreams, and the process of thinking, and inferred that thinking is the result of superposition of multiple types of knowledge. The hypothesis they proposed organically integrates unsolved biological issues, such as the sleep process, the mechanism of dreams,

the storage of thoughts, and the processing of memory information, and can be employed to explain various phenomena in the cognitive thinking process. It provides a new and exploratory way for finally unveiling the mystery of the thinking mechanism. In 2007, Jeff Hawkins [28] described the relationship between machine and intelligence. These descriptions are similar to those that are currently used in deep learning.

The lack of understanding of the working mechanisms of human brain is one of the biggest limitations for the development of intelligent machines. Neuroscience has not yet revealed the operating mechanism of human brain, which constitutes a constraint on this research. Because of this, most of the AI algorithms that are currently utilized, are only based on a statistical interpretation of neuroscience, and a focus in this area is the development of new algorithms that mimic actual human brain.

AI is a very powerful and exciting area. It will become more important and universal in future, and will have a major impact on modern society. ANNs and more sophisticated deep learning techniques are the most advanced and effective AI tools for solving very complex problems. Although deep learning sounds elegant, it is essentially just a term used to depict a particular type of neural network and related algorithms. In these algorithms, raw input data are typically processed by multiple layers of nonlinear transforms for feature extraction. Unsupervised feature extraction is an excellent part for deep learning. It refers to a meaningful feature of the data that is automatically extracted through an algorithm for further learning, generalization, and understanding. Feature extraction usually involves a certain amount of dimensionality reduction, it reduces the dimension of input features and produces reasonable results. It promotes simplification, computation, memory and power reduction. Layer number of deep learning model is larger than that of shallow learning algorithm. Shallow learning is often less complex, compared to deep learning, it requires more prior knowledge of the use of optimal features, and often involves feature extraction in other machine learning methods, as well as cumbersome feature selection and engineering.

More generally, deep learning belongs to feature learning or representation learning techniques. Feature extraction is used to “learn” features of interest. Feature learning algorithms allow machines to learn specific tasks using a set of features. Deep learning gives computer system the ability to “learn” features from data.

It integrates representativeness, goals, and optimization to progressively improve the performance of specific tasks. Performance is crucial for machine learning. An efficient representation learning can effectively transfer raw data to a machine learning system. In addition to better hardware, there are two other factors for the success of deep learning: 1) deep neural network, which enables a more efficient learning from a large amount of data, compared to traditional methods; 2) end-to-end (supervised) deep learning, providing the ability to learn direct mapping from input to output.

AI and computer vision has made significant impact on human society and will contribute more in future. Computer vision describes the machine's ability to process and understand visual data; it automates the type of tasks that human eye can accomplish. In layman's terms, computer vision is the application of AI to the visual world.

One of the most important breakthroughs in AI was made in 2011. Before this, scientists thought that for a machine to think and interact like a human, uploading of large amounts of knowledge/data to the computer is required, a lot of rules needs to be defined for how to process the data. However, in 2011, researchers in the AI team of Google built a wonder, using deep learning. Instead of entering the rules themselves, they provide the computer with a large amount of accurate labeled data and let the machine itself analyze the data - this process is called "supervised learning". The machine can then identify unlabeled data that was fed to it, and learn how to analyze it from ingested labeled data.

This breakthrough has led to AI-enabled personal digital assistants, such as Siri, Google Translate and Google Assistant, which are widely used. AI technology has been successfully utilized in industry and society, including fields of health care, transportation, city, etc. In medical field, image recognition is increasingly used to analyze X-rays, MRI and other scans. Autonomous vehicle is another example for application of AI - e.g. the AI system "Tesla Autopilot" enables autonomous drive and navigation, using the camera and other advanced sensors equipped on a Tesla car. Similarly, computer vision has been used in "smart cities" to solve traffic and crime problems; it will be used in many market applications. AI and computer vision are expected to become an integral part of almost all parts of industry in near future, and have received intensive interests from researchers all over the world, which motivates the current study.

2.2.1 Region Proposal

Humans can detect visual distinction, so-called salient (ie, pre-attentive stage), effortlessly and quickly. Being able to perceive an object before it is recognized, is related to bottom-up visual attention. Under the definition of saliency, related research is divided into three categories: fixed forecasting [29], salient object detection [30] and objectness proposal generation [31–33], which has evolved considerably in the last two decades, especially since 2007 [2, 34].

Fixed prediction model aims to predict the significant point of human eye movements [35]. Despite significant developments in gaze point prediction models, the model prediction tends to highlight edges/corners, instead of the whole object. Therefore, these models are not fit for regional proposal generation. In contrast, the salient object detection model attempts to detect the most striking object in the scene and then segment the entire object range [36]. In the past two decades, a large amount of work has been put forward for salient object detection. Except for some models that try to segment the object of interest (e.g., [37, 38]), the goal of most of existing models for salient object detection is to first identify significant subsets (i.e., compute a saliency maps) from an image, and then integrate them to split the entire salient objects. In [39], Itti et al. employed a few examples to demonstrate the ability of their models in detecting spatial discontinuities in a scene. Since then, a lot of salient object detection models have been proposed and exploited.

For region-based models with intrinsic cues, three types of algorithms are mainly used for saliency computation, i.e. mean shift algorithm [40], graph-based segmentation algorithm [41] and clustering (quantization); these algorithms can be used to generate irregular regions of different sizes. On the other hand, with recent advances in super-pixel algorithms, compact areas with comparable size are also popular choices, through the use of simple linear iterative clustering (SLIC) algorithm [42], the turbo-pixel algorithm [43], etc. Uniqueness is the most commonly used feature for measurement of saliency of a region. Models with external cues use external cues to help detect prominent objects in images and video. However, saliency annotations are time consuming, cumbersome, and difficult to collect; hence, these methods are probably not suitable for complex images with many objects. A recent interest in this area is the recovery of convolutional neural network (CNN), especially the introduction of FCN in [44]. The CNN-based approach

FIGURE 2.2: An overview of salient object detection [2]

eliminates the need for hand-crafted features and reduces the reliance on central bias knowledge, it is therefore used by many researchers. In general, an overview of salient object detection is presented in Figure 2.2.

Due to its ability to learn multiple levels of features, CNN can easily and accurately position salient objects. Low-level features include the edges that constitute the boundaries of salient objects, while advanced features allow the identification of salient objects in conjunction with semantic information. Various CNN-based models have been developed recently. In these methods, high-level features are usually gradually propagate back to lower layers, to allow efficient fusion of multi-level features; another way is to update a more powerful baseline model, e.g. using ResNets [45], instead of VGGNet [46]. Last few years, object detection has made great progress and has become one of the most important areas of computer vision. However, region of interest (ROI) proposal remains a bottleneck for target detection. To reduce the number of ROIs, measurement of objectness have become popular recently [47–52].

FIGURE 2.3: Image labeler to set ground-truth

For faster region-based CNNs (faster R-CNN)[53–55], images are labeled with objects as ground truth for supervised learning. Figure 2.3 demonstrates an example for ROI labeler, whereby the features of the labeled/cropped image are extracted for feature learning, classification and model generation. In the test, approximately 300 randomly generated bounding boxes are provided for each image as ROI proposals, each ROI is predicted by a pre-trained model. Intersection of union (IoU) is applied to compare the difference between detected object and the ground truth, and the objects with higher IoU scores are considered as the detected objects. Linear regression and linear classifiers are applied during training and prediction. The objects detected in the test are close to the ground truth, indicating good detection accuracy. Despite this, the labeling work of ground truth is time consuming, tedious and inefficient, as more and more data sets require repeating and endless labeling work.

Considering the labeling issue of faster R-CNN, “You only look once” (YOLO)[3, 4] is applied to propose detected objects for ROI proposal in this study. YOLO is a state-of-the-art real-time object detection system. Faster R-CNN recommends about 300 bounding boxes as ROI proposals for each image. The amount of objects which is proposed by YOLO is much less than faster R-CNN. The key is that YOLO divides the input image into 7×7 units. Each grid unit only predicts one

object through a fixed number of anchor boxes. It predicts twenty conditional class probabilities. The PASCAL Visual Object Classes (VOC) 2012 dataset [56–58] is used as a reference for objectness detection, and square root in loss function is used for regression and detection. The PASCAL VOC Challenge is a very popular data set for building and evaluating algorithms for image classification, object detection and segmentation. The Pascal VOC 2012 dataset has 20 classes. The train/validation data have 11,530 images containing 27,450 ROI annotation objects and 6,929 segments.

Objectness is usually represented by values that reflect the likelihood that an image window will cover an object of class [35]. It can be used in ROI proposal to increase computational efficiency; since it has good generalization capabilities for invisible object classes, these ROIs can be reused by other specific categories of detectors, to reduce the computational cost. In [31], binary normed gradient (BING) was used for object estimation, it requires only a small number of atomic operations; its detection performance is usually better than other detection approaches, and could be more than 1000 times faster, compared to some popular methods [48, 49, 51]. In addition, in order to achieve efficient sliding window object detection, the feasibility of maintaining computational cost is very important [59, 60]. The concept of objectness was extended in [3, 4].

2.2.2 Object Classification

Object classification is easy for human but complicated for machines. Classification includes image sensing, data preprocessing, feature extraction, training and object prediction. Many classification techniques have been developed for image classification, e.g. [61–63]. Image classification is an challenging task in various applications, it is complicated and relies on many factors. A discussion on the current technology, issues and prospects for image classification was presented in [62]. The main concern of this investigation is advanced classification technology.

Images gained from different sensors have unique features. Data fusion could significantly improve visual interpretation and quantitative analysis. It is preformed in three levels, i.e. pixels, features, and decisions, and involves two main steps:

FIGURE 2.4: ImageNet classification challenge

(1) geometric registration of two data sets; (2) combining spectral and spatial information content to generate a new data set that contains information from two data sets.

ImageNet [64] project is a large-scale visual database dedicated to the study of visual object recognition, it is developed by Professor Li Feifei and her team from Stanford University. Since 2010, the ImageNet project runs an annual software competition, using a “trimmed” list of 1,000 non-overlapping classes. The deep learning revolution in 2010s leads to a huge breakthrough [65] in solving the ImageNet challenge in 2012: “Suddenly, people began to pay attention not only to the AI community but also to the entire technology industry” [66]. The root cause is that CNN is used for extraction of features that include spatial information content, which is different from other traditional features. In 2015, three years after the release of ImageNet, the computer has shown better performance than humans in identifying objects [67]. As shown in Figure 2.4, the solid blue line shows error rate and the orange dotted line shows the revolution of CNN layer depths.

For deep learning, the biggest challenge is that it requires optimization of many parameters, which usually results in high costs in time and machine resources. In the era of big data, data fixation is impossible, so is model fixation, because the

dimensionality of data needs to be adjusted from time to time, such as adding new features. Instead, incremental learning algorithms (e.g. broad learning) were employed.

Deep learning deepens the network layer, increases the complexity of the model, and better approximates the nonlinear functions we want to learn, but nonlinear layers are not as much as better. This theory has long proven that a single layer feedforward network (SLFN) can be used as a function approximation. It indicates that there is no need to increase the number of layers. RVFLNN has also been shown to be able to approximate any continuous function on a compact set. Its nonlinear approximation ability is reflected in the nonlinear activation function of the enhancement layer.

2.2.3 Model Optimization and Fine-tuning

It is well known that for effective feature learning, low bias and low data variance are preferred. If not available, then at least low bias and high variance are preferred as shown in Figure 2.5(a).

Errors with any model can be divided into three parts: deviation, variance, and irreducible errors. High bias error means that the model performs poorly and constantly ignores key features (under-fitting). Variance measures how the predictions for the same observation differ from each other. The high variance model will over-fitting the training samples and will not perform well on the test samples. In Figure 2.5(b), it interprets high bias as “modeled data is insufficient” and high

FIGURE 2.5: Bias variance tradeoff

variance as “excessive modeled noise”, in other words, the complexity of the model is too low to capture meaningful data or the complexity is too large and captures lots of noise too. To improve model performance, the complexity of the model is increased, and the error is reduced due to lower bias in the model, but it has been working to a certain extent. As the model continues to be more complex, it will eventually over-fitting, so the model will begin to suffer from high variations. A better model should balance between bias and variance errors. Therefore, validation patience is added to perform this trade-off analysis. Validation patience value is the number of times that the loss on the validation set can be larger than or equal to the previously smallest loss before network training stops. It is an early stopping to halt the training of neural networks at the right time.

The related technologies in fine-tuning [68, 69] are beginning to be popular in different fields, such as image classification, natural language processing, including CNN/recurrent neural networks(RNN) layers’ fine-tuning and classifier’s fine-tuning.

For limited data, data augmentation through balanced dataset or continuous generative adversarial network (GAN), can be applied, to improve the generalization of the system.

2.2.4 Object Detection

Deep learning based object detection starts from Two-stage deep learning object detector, R-CNN [54] in 2014. Typically, in R-CNN network, the region proposed network (RPN) [51, 53] is used as an external module, independent of the detector. The region proposal method includes a sliding window method (for example, objects in the window [48] and border [70]), grouping super-pixel method [71] and regionlet for general object detection.

R-CNN uses selective search to improve search speed, but it is still relatively slow because it passes through the proposed ConvNet for each object without shared computation. Spatial pyramid pooling (SPPnets) [72] is developed to accelerate R-CNN by sharing computation. In terms of test speed, it accelerates R-CNN by a factor of ten to one hundred, but its training is a multi-stage pipeline. In contrast with SPPnet, fast R-CNN takes an overall image and a set of ROIs as input. It overcomes the shortcomings of SPPnet and R-CNN, allows faster training/testing;

its training uses single-stage multitasking loss and has higher detection accuracy. From the comparison among regional proposal models in [73], edge-boxes, multi-scale combinatorial grouping (MCG) and selective search have a better detection rate. The edge-boxes method shows the best balance between implementation and speed. Despite all the efforts to improve speed, regional proposals still take a long time. Faster R-CNN shares the convolution feature with the “attention” mechanism, and combines RPN and fast R-CNN into a unified network. That is, the RPN proposes region of interests to fast R-CNN detector for prediction. Faster R-CNN uses singular value decomposition (SVD) to increase learning speed. It learns through RPN and FCN. As an advanced CNN network, faster R-CNN can reduce runtime and achieve near real-time object detection. In this thesis, deep learning model was used to extract full convolution activation feature (CAF), and perform transfer learning of the CAFs for image-based defect assessment of buildings.

In faster R-CNN framework, ROIs are manually labeled; for different anchors, IoU values are calculated according to comparison between anchors and ground truth boxes, and linear regression and non maximum suppression (NMS) are employed to find optimal boundary for different objects. Labeling process is very time consuming for new category objects. Computing for ROIs using the method in faster R-CNN is still time consuming; hence, $N \times N$ unit segmentation method of an image and objectness concept in YOLO [3, 4] are preferred, because its detection performance is much better than the ROI proposal method in faster R-CNN, and could be more than 1,000 times faster than some of the most popular methods [48, 49, 51].

In 2016, Single-stage object detection network, You only see once (YOLO) is developed using $S \times S$ grid and each grid’s anchor bboxes for ROI proposal. YOLO v2 object detector uses a single-stage object detection network. YOLO v2 is faster than other two-stage deep learning object detectors, such as faster R-CNN. YOLO v2 uses anchor boxes to detect object classes in the image. For more details, see the anchor box for object detection. YOLO v2 predicts these three properties for each anchor box:

- Intersection over union (IoU) - predicts objectness score of each anchor box.
- Anchor box offset - Optimize anchor box position.

FIGURE 2.6: Bounding box object detectors: understanding YOLO [3, 4]

- Class probability - Predicts the class label assigned to each anchor box.

Figure 2.6 (b) shows the predefined anchor box and the refined position after applying the offset (thick lines). To help understanding, an reinterpretation of the characteristics of the anchor box object detector in YOLO is depicted as follows. Convolution enables the prediction of different locations in an image to be performed in an optimized manner. This avoids the use of sliding windows to calculate the prediction of each potential location separately. For positions on the unit, the one that is closest to the ground truth, is positive; other positions are negative. The cells in Figure 2.6 (a) collect all possible locations in the center of the ground truth box to activate the network output as positive. For each positive position, the network predicts the regression of the exact position and size of the anchor box. In [4], these predictions are related to unit position and anchor size (rather than full images) for better performance as in faster RCNN. Figure 2.6 (b) indicates that (c_x, c_y) is the grid unit coordinate and (p_w, p_h) is the anchor size.

$$b_x = \sigma(t_x) + c_x$$

$$b_y = \sigma(t_y) + c_y$$

$$b_w = p_w e^{t_w}$$

$$b_h = p_h e^{t_h}$$

After the bounding box regression is trained, the model is also trained to use the regression model described above to predict the confidence score on the bounding box of the final prediction. Confidence score is the IoU score between predicted image cell and ground truth. YOLO network has two major component: CNN feature extractor and two fully connected (fc) layers. The convolutional part is adapted from Googlenet. The Input is $3 \times 448 \times 448 (D \times W \times H)$, and CNN output is $1024 \times 7 \times 7$. For an example of 20 object classes, B is used to represent positive or negative confidence scores, C stands for 20 object classes, and S means to equally split the image into $S \times S$ parts. After flattening the dimension to 50176, these features will pass through two fc layers, feature dimension is then decreased from 50176 to 4096 followed by $(B \times 5 + C) \times S \times S = 1470$ (e.g., $S = 7, B = 2, C = 20$). In the end, reshape the dimension to $30 \times 7 \times 7$. As shown in Figure 2.6 (C), YOLO outputs 1470 vectors. YOLO can not directly conduct transfer learning, but the $S \times S$ boxes concept and object confidence scores can be referred. The predicted bounding boxes for objectness detection, together with new category class prediction, can transfer learning skills to new object detection. For active transfer learning for convolution activation features (A-TLCAF) and online TLCAF, objectness is used as the first step for region of interests' proposal in this study.

2.3 Transfer Learning

Transfer learning (TL) aims to improve; but in certain situations, transfer learning could have the opposite effect, which is referred as negative transfer. For a learning task, only features that are applicable to all individuals should be transferred; if a feature is unique to certain individuals and not applicable to others, its transfer could lead to negative transfer and should be forbidden. In some cases, a negative transfer may occur when the tasks of the target domain and the source domain are not related to each other. In transfer learning, it is supposed that different tasks are relevant. However, definition and mathematical description of the correlation between different tasks are subjective decisions that are biased toward human. For transfer learning, there are four requirements: 1) there should be relationships

between tasks; 2) these relationships can be obtained in a computational way without the input of knowledge/information from human; 3) various tasks belong to a structured space, and are not independent concepts; 4) a unified model is provided for transfer learning. For tasks selected, questions are raised, such as: are there relationships and how many relationships between them? To answer these questions, a global understanding of the relationships and redundancy between tasks is required. For this, the tasks should be treated as a group, and not individually. The relationship and redundancy between tasks are used to improve efficiency. One of the most interesting parts to increase is the efficiency of supervision, which is to solve problems with less labeled data. This is main issue of this work. Numerous research papers have discussed how to reduce model dependence on labeled data, and approaches proposed include self-supervised learning, unsupervised learning, meta-learning, task adaptation, and fine-tuning based on the features learned on ImageNet. Nowadays, transfer learning has become a common practice. Transfer learning depends on the relationships between tasks. At a high level of abstraction, if the internal state of a task could be learned by a model, transfer and translation of it may help solve another task if there is a relationship between the two tasks. ImageNet is frequently employed as a pre-trained model for fine-tuning, because the large data set of ImageNet itself guarantees a reliable generalization of the learning network. However, what if there is only one small data set available? Will the result be the same? Or will it be even worse than learning from zero? For example, DeepMind's new AlphaGo Zero which learns from zero is much better than the previous version of AlphaGo Lee based on human atlas.

Reusing existing knowledge domain data, a large amount of existing work will not be completely abandoned. There is no need to pay a huge price to reacquire and calibrate a large new data set, and it is also possible that the data will not be available at all. For rapid emergence of new areas, they can be quickly migrated and applied, reflecting time-saving advantages.

Baidu's former chief scientist, Professor Andrew Ng, said in his popular Neural Information Processing Systems (NIPS) 2016 tutorial that "transfer learning will be the next driving force behind machine learning's commercial success after supervising learning". As shown in Figure 2.7, there is no doubt that supervised learning is currently most mature. It can be said that supervised learning has been successfully commercialized and the next commercial technology will be transfer learning.

FIGURE 2.7: Driving force behind the success of machine learning industry[5]

In short, transfer learning will become the next exciting research direction. As the name implies, transfer learning involves transferring pre-trained model parameters into new models to help the new models train data set.

The success of ML in industry is largely driven by supervised learning. Inspired by deep learning, more powerful computing utilities and massive datasets have spurred a wave of renewed interests in AI. In recent years, this has become part of our daily lives. Transfer learning not only understands the features learned from existing data, but also provides knowledge about the nature and methods of neural network “learning”. Recently, the number of articles and research on transfer learning has been constantly increasing, but there are still many open areas/issues for exploration. It is hoped that the results of this study can give readers a certain degree of awareness of transfer learning.

Transfer learning will be the next exciting research direction. In particular, many applications require models to be able to transfer knowledge to new domains. When source domain and target domain have some commonalities (e.g., overlapping features), the transfer learning method can identify the knowledge learned from the source domain and adapt it to the target domain [6, 74]. Feature-based transfer learning algorithms learn transformation by encoding high-level features. Advanced domain-specific features are able to help locate learning tasks, even if only limited labeled data are available.

FIGURE 2.8: Traditional supervised learning settings in ML

As the name implies, transfer learning involves transferring pre-trained model parameters into new models to help them training data set. In contrast, in traditional supervised learning setting, the labelled data are given to train a model for a task,, as shown in Figure 2.8, whereby the training data and testing data come from the same domain.

For instance, to identify defects in an image (e.g. those come from Building & Construction Authority (BCA)’s CONQUAS Room), for transfer learning, a model can be trained using this dataset (the image), the model can then be applied to unseen data for the same task and expected to perform well. While for supervised learning, the model trained using existing data is generally not applicable to new data (not in the training dataset of the model), labeled data in new domains are required to train new models to achieve reasonable accuracy. Moreover, traditional supervised learning paradigm collapses when there is not enough labeled data for the model to train. In practice, performance degradation is often experienced for supervised learning, because the model inherits its training data’s bias and has no idea on how to generalize to new domain. If the model is trained for new tasks, the existing model can not be reused, since new labels for new tasks are different from old labels for the existing model.

Transfer learning is able to handle these scenarios by leveraging existing labeled data for certain related tasks. The knowledge gained in solving the source task in the source domain is able to be stored and used to the new problem we are

FIGURE 2.9: Transfer learning settings

interested in. The learning process of transfer learning is illustrated in Figure 2.9. In fact, transfer learning is required to transmit as much knowledge as possible from source domain to target domain. The knowledge could be in various forms depending on the data, e.g. it can be about the general shape and texture of objects.

In transfer learning, the underlying network on the underlying data sets is trained first, the learned features are then transferred to the second target network to train the target data sets. If these features are generic, then this process tends to work, which means it works for both basic and target tasks. Data driven machine learning methods aim to develop a algorithm that can automatically improve its performance through experience. It is the core of AI and data science, and one of the fastest growing areas of technology today. It enables evidence-based decisions in many industries. The main issues of transfer learning include when to transfer, what to transfer, and how to transfer. Most of the current papers focus on what to transfer and how to transfer, and research on when to transfer seems limited, although it is also a key factor in transfer learning.

It is assumed that the data used follow the same distribution and are derived from the same feature space during training and reasoning. However, in practical applications, it is difficult to rigorously fulfill this assumption. Some problems that are frequently encountered include:

1. Limited labeled training samples - for example, when processing classification in the target domain, there are not enough training samples; at the same time, there is a large amount of training data related to the target domain in the source domain, but the source domain and the target domain are located in different feature spaces, or samples from source and target domain follow a different distribution.
2. Variation in data distribution - data distribution may change with time, location or other dynamic factors. The data collected previously are probably outdated; under these circumstances, it is necessary to recollect the data and rebuild the model.

Transfer learning is an optimization technique in which a trained task model is reused for a second related task. When modeling the second task, it can achieve rapid progress or improved performance. In general, transfer learning provides an advantage if there is enough data in the source domain but there is less data in the target domain. For example, for task classification, the data in the target domain are limited, but there is a large amount of relevant training data in the source domain; the data characteristics distribution of the source domain and the target domain are different. In these cases, transfer learning can be used to improve the recognition rate of under-sampling tasks.

So-called transfer learning is typically used to transfer knowledge gained from the source domain to the target domain. There is usually a domain difference between the source domain and the target domain, and the data in the source domain and the target domain can follow different distributions. Most importantly, if the “posture” of transfer learning is wrong, it will probably result in negative transfer, which means using the target domain alone is better than adding data from the source domain.

TL can be classified based on the content transferred [6], as illustrated in Figure 2.10.

1. Instance-based TL - it weighs the importance of the samples, and gives more credit to samples with higher importance.
2. Feature-based TL - it extracts a good representation of the feature from the data in the source domain, then encodes the knowledge in the form of features and

FIGURE 2.10: TL classification (based on content transferred)

passes it from the source domain to the target domain to improve the performance of the task in the target domain.

3. Model-based TL - in this TL approach, tasks in target and source domain share the same model parameters or follow the same previous distribution. The entire model is applied to target domain, such as fine-tuning of the pre-trained deep network. It is also called parameter transfer.

4. Relational transfer learning - in this TL approach, establish a mapping of the relationship between source domain and target domain, and assume that both the source domain and the target domain are relational domains.

For the first three TL methods, it is assumed that the data is independent and follows the same distribution. Moreover, all four types of transfer learning methods require the source domain to be related to the target domain.

According to whether the source domain and the target domain are the same, whether they have labeled data, transfer learning can also be classified as shown in Figure 2.11. The form of transfer learning used in deep learning is called inductive transfer. It can be viewed as a directional search in a given hypothesis space [75], as shown in Figure 2.12. Inductive transfer uses source task knowledge to adjust for induced bias, including using a model suitable for different but related target tasks to narrow the hypothesis space or search steps in a beneficial manner.

FIGURE 2.11: Overview of the different settings of transfer learning [6]

FIGURE 2.12: Illustration of inductive transfer

2.4 Online Learning

In actual databases, the amount of data tends to increase, so learning methods should be able to further train the model to extend the knowledge of existing models to incorporate new input data. Online and incremental learning algorithms are able to adapt to new data without forgetting the existing knowledge. It does not retrain the model. Some incremental learning algorithms have built-in parameters to control the correlation with old data and learn the representation of training data without partially forgetting the old data over time.

Online learning is a collection of machine learning methods which learn from sequential data. The online learning model is constantly updated as new data arrive. Online algorithm is a growing but often misunderstood branch of machine learning, in which model parameter estimates are updated for each new piece of information

received. While small batch methods are often incorrectly labeled as “streaming machine learning”, true online methods have different implementation and goals. This section describes the main differences between online and offline machine learning, as shown in Table 2.1, and introduces common online algorithms, and how to analyze online algorithms.

Due to its application in the era of big data, online learning has gained significant popularity and attracts extensive research interests. The advantages of online learning’s three-V feature is introduced as follows.

Volume: Online learning requires smaller data storage. For huge data, it is often impossible to read all of the data at once. Since online learning uses only the most recent data points to update its model, the system does not need to keep plentiful data in memory. Systems using online learning maintain much less data storage than those based on offline learning. The ability of online learning to generate a model from streaming data, provides benefits to many applications, especially those with high throughput and limited memory.

Online Input processed piece by piece in a serial fashion, model is updated as new data arrive, and previous data is not stored. The model gradually improves as it receives more data points. It can be all four types of learning: supervised, unsupervised, semi-supervised and reinforcement learning.

Offline It treats data as a static pool, and assumes that all data are available at training time. It produces only one final model for all data at once. It Can be supervised, unsupervised or semi-supervised.

Online Incoming data →current data (used to train and update the model)
→past data →dispose

Offline All data accumulated →model generation

Online Each new message will generate an event. Data arrive in order and the observation window is limited. For instance, online video processing. Online learning can follow an active learning strategy, and the algorithm queries the data in order. E.g., interactive machine vision assists robot inspection.

Offline Input processed in batches.

TABLE 2.1: Differences between offline and online learning

Velocity: Online learning allows quick model updates. The design of online learning takes speed into account. By processing only a small block data at a time, the online learning method maintains a relatively low computational complexity for each update. In contrast, retraining offline models using offline learning method, can be extremely costly when frequent update of model is required. Therefore, the online learning method is better than the offline one, and is a natural choice for applications with tight time budgets, such as real-time machine vision processing that requires a response time of about a few milliseconds.

Variety: online learning adapts better to changes in data. Many online learning methods have “forgetting factors” that allow the user to set a speed at which the methods forget the past. Online learning achieves this by gradually discounting the importance of past data. As a result, the model automatically adapts to changes in the dataset. Meanwhile, sudden changes in the model also indicate changes in the data. On the contrary, offline models are not suitable for change detection, because they are not updated frequently and can result in significant delay in detection.

Table 2.2 below presents a comparison among a wide range of online learning algorithms, including model-dependent methods (e.g. incremental support vector machines), and model-independent methods (e.g. stochastic gradient descent). Most of the online learning algorithms respond to non-stationary environments by introducing a forgetting mechanism, but the major problem in this research area is progressive learning in a stationary environment, whereby forgetting mechanism is very harmful and reduces performance.

Method	Incremental Support Vector Machine (ISVM)
Benefit	Incremental version of SVM [76] is a lossless algorithm that produces the same model as the corresponding batch algorithm [77, 78]. The online approximation SVM solver was proposed in [79] and applied to [80, 81].
Limit	It is complex, slow to train and run, and does not support endless learning.

Method Online Random Forest (ORF)

Benefit Incremental random forest [82]. The tree collections are popular because of high precision, simplicity, parallelism and not sensitive to feature scaling [83].

Limit It does not support endless learning or concept drift.

Method Naive Bayes (NB_{Gauss})

Benefit It applies to the axis parallel Gaussian distribution of each class and is used as likelihood estimates [84]. The sparse model allows efficient learning considering processing time and memory requirements. It can effectively learn from a small number of data [85, 86].

Limit Independence assumption for features and the inability to handle multi-modal distribution.

Method Incremental Extreme Learning Machine (IELM)

Benefit It re-forms the batch Extreme Learning Machine (ELM) least squares solution into a sequential solution [87]. It can handle data one by one or in chunks efficiently.

Limit Compared to hidden neurons used, more examples are required for effective initialization of output weights [88, 89].

Method Stochastic Gradient Descent (SGD)

Benefit It is an effective optimization method to learn the discriminant model by minimizing the loss function, such as hinge or logistic loss. It can be combined with linear models, and is especially effective for sparse high-dimensional data [90, 91].

Limit Linear models do not provide good results, whenever a nonlinear class boundary is required, which is common for low-dimensional data.

Method Incremental Learning Vector Quantization (ILVQ)

Benefit A dynamically growing model [89] inserts a new prototype when necessary, and the insertion rate follows the number of samples misclassified.

Limit Training and running speed is not the fastest.

Method	Fuzzy Adaptive Resonance Theory (Fuzzy ART)
Benefit	Maximize the acceptance of new knowledge (flexibility) while ensuring less impact on past pattern samples (stability) [92–94].
Limit	It is unsupervised learning
Method	Deep Q-Learning network (DQN)
Benefit	It is a type of deep reinforcement learning for action control, use deep neural networks for generalize and pattern match between states instead of a lookup table. It uses experience replay to avoid the instability of the nonlinear function approximator [95, 95, 96, 96, 97].
Limit	Its target is non-stationary or unstable.
Method	Random vector functional-link neural networks (RVFLNN)
Benefit	Fast and dynamic learning. It eliminates the shortcomings of the long training [98].
Limit	It uses raw data as an input layer.
Method	Broad learning
Benefit	It is extended based on random vector functional-link neural networks (RVFLNN). Compared to RVFLNN, broad learning performs some conversion of the data, which is equivalent to feature extraction. It supports the expansion of multiple feature nodes, as well as enhanced nodes. Besides, it supports new incoming input data. It fuses heterogeneous signals for synergistic knowledge discovery [99].
Limit	Current experiments is focused on deep learning for perception parts

TABLE 2.2: Comparison of online learning algorithms

An analysis on the most common incremental learning algorithms for different, static and non-static data sets, was performed in [100]. ISVM provides the best precision at the expense of the highest complexity; ORF has a poorer performance, but a very fast training and running speed. The disadvantage of the ORF model and the ISVM model is that they all follow a linear increase in the number of samples and cannot be controlled in a straightforward manner. So, they are not suitable for learning in an infinite stream, with easily constrained complexity. ILVQ

provides an accurate and sparse alternative to ISVM, and its tree-based model is especially suitable for high-dimensional data due to its compressed representation and sub-linear runtime, regardless of dimension, but the compression representation violates the learning speed, making the instance-based model fast convergence and more suitable for tasks that include only small samples. Sparse models like SGD and NBGauss are suitable for large sample learning in high dimensional space and are not complicated for low dimensional tasks.

Deep learning requires optimization of a large number of parameters, this is the biggest challenge/difficulty for deep learning, it takes a lot time and machine resources to do.

In fact, there are no perfect algorithms, each has its own advantages and disadvantages. The real world comes from a complex environment that develops over time, this imposes a big challenge for learning algorithms. For online DTL, it focuses on designing algorithms that deal with concept drift; however, the changes in the underlying distribution of data could lead to inaccuracies over time. Online learning algorithms face many challenges. Many online learning methods require optimized initialization to avoid long convergence times, and frequent model updates can result in unnecessary fluctuations in model output. Despite these challenges, on-line learning has become more and more prevailing, because of its high-speed data processing capabilities and ability to quickly adapt to data changes.

Since DeepMind pioneered a deep reinforcement learning (DRL) model to play Atari games [96], DRL has become a common method that enables agents to learn complex control strategies in various video games. [95] used a convolutional neural network to train the variants of Q learning, designed the reward mechanism, and introduced a dual experience playback method to classify different experiences to obtain better training effects. It shows that their agents are better than the baseline model in snake playing games and outperform human level performance. [97] introduced a new approach to computer Go, using the Value Network to evaluate board positions and policy networks to select actions. [97] also introduced a new search algorithm that combines Monte Carlo simulation with a value and policy network. Using this search algorithm, their program AlphaGo and other Go programs have a winning rate of 99.8%. This is the first time a computer program has defeated a human professional player. In the literature review of the deep transfer learning domain, it is noted that the deep Q-learning network (DQN) is used as a

DRL for action control, using deep neural networks for generalization and pattern matching between states instead of a lookup table. It uses empirical replay to avoid the instability of nonlinear function approximators. The limit is that the goal is non-stationary or unstable.

Another method, broad learning [99], is extended based on the random vector functional-link neural networks (RVFLNN) [98]. RVFLNN is a fast and dynamic learning network. It overcomes the shortcomings of the long training process, but it uses raw data as the input layer. Compared to RVFLNN, broad learning performs some conversion of the data, which is equivalent to feature extraction. It supports the expansion of multiple functional nodes, as well as enhanced nodes. In addition, it supports new input data. It combines heterogeneous signals for synergistic knowledge discovery. Therefore, an online-TLCAF model is proposed in this thesis. Online-TLCAF uses machine learning assembly method and deep transfer learning method to build an automated object detection model. The model can automatically propose potential objects, use CNN to transfer low-level object features, and then perform feature extraction and online incremental learning on different types of objects until object detection. These facilitate a quick learning of the high-level CNN features of different objects, and could adapt to the changes of new data, without forgetting old experiences. Current experiments focus on deep learning in the perceptual part.

2.5 Summary

There is a proverb standing on the shoulders of giants. This usually means that people should be good at learning the experience from their predecessors. Transfer learning has the same meaning as this phrase in the environment of machine learning. When processing a task, it is always helpful to learn from the existing resources. Therefore, an embodiment of deep transfer learning, called active transfer learning for convolution activation feature (A-TLCAF), and an online deep transfer learning method (named as online-TLCAF) are proposed in this report, this will be described in Section 3.

The future AI will be able to understand us, interact with us, collaborate with us, and even augment with us in human ways. AI is no longer a science in laboratories,

it has become one of the biggest driving forces for the fourth industry revolution. It will have a profound impact on how human beings live, work, and will shape our environment. What role should such a powerful technology play in our world? How human will be affected? Like any other technology, AI is just a tool in the hands of human. It is therefore believed that there is no independent machine value, and machine value essentially comes from human value. AI will change the world and its impact on the society is kept growing. Therefore, it is important to study AI and use it in a proper manner to benefit human society. There are still many areas in AI remain to be explored, which motivates this investigation.

Chapter 3

Online Deep Transfer Learning

Supervised learning tasks (such as classification) often require good input representations to achieve superior performance. In general, pre-training model does not guide the meaningful features of building defects and requires the use of TLCAF to obtain linearly differentiated features. In this chapter, to process data at high speed and quickly adapt to data changes, an online deep transfer learning (DTL) methodology, which is called online transfer learning for convolution activation features (online-TLCAF) is presented, as shown in Figure 3.1. First, the modified YOLO network is used for objectness proposals to transfer object features from the source domain to the target domain. After that, TLCAF is applied to adapt the pre-training network to the target domain for feature extraction. The features are then extracted by TLCAF and passed to broad learning of online incremental learning. Finally, the predicted objects with their positional information are reconstructed in the original image for object detection and visualization. Details of the procedure are described in Section 3.3.

Deep learning attempts to learn advanced features from a large amount of data, it makes deep learning beyond traditional machine learning. It can automatically extract functionality in an unsupervised way. In contrast, traditional machine learning requires manual feature design capabilities, which increases user's load. Deep learning is a representation learning algorithm based on large-scale data, in which data dependence is one of the most serious problems. Compared to traditional machine learning methods, deep learning relies heavily on a large amount of training data since it requires massive data to understand the underlying data

FIGURE 3.1: Online-TLCAF

patterns. In some special areas, insufficient training data is an inevitable problem that poses a challenge to deep learning.

Transfer learning helps solve the basic problem of insufficient training data in machine learning. It attempts to transfer knowledge from the source domain to the target domain by relaxing the assumption that the training data and test data must be independently and equally distributed. In transfer learning, there is no need to train the model in the target domain, which significantly reduces the need to train data and decreases training time in the target domain. This section reviews the latest research work on deep transfer learning technology [7].

Transfer learning's learning process is shown in Figure 2.9. The domain and the task are defined first. The domain is indicated by $D = \{\chi, P(X)\}$, it comprises two portions: the feature space χ and the edge probability distribution $P(X)$, where $X = \{x_1, \dots, x_n\} \in \chi$. A task is expressed by $T = \{y, f(x)\}$, it includes two parts: the label space y and the target prediction function $f(x)$; $f(x)$ can also be considered as a conditional probability function $P(y|x)$. Then, transfer learning can be defined below.

Definition 1. Transfer learning

For a learning target task T_t based on target domain D_t , help can be gotten from source domain D_s for a learning source task T_s . Transfer learning aims to increase the performance of prediction function $f_T(\bullet)$ for T_t by discovering and transferring

FIGURE 3.2: Categorization of DTL [7]

latent knowledge from D_s and T_s , where $D_s \neq D_t$ and/or $T_s \neq T_t$. Additionally, in most circumstances, D_s size is much bigger than D_t size, i.e. $N_s \gg N_t$ [7].

The existing research on transfer learning was summarized in [6, 101], which introduces many classic transfer learning methods. In recent years, deep learning has gained more and more applications in various applications. An important aspect of deep learning is how to effectively transfer knowledge through deep neural networks (DNN), which leads to the development of DTL methods, which are defined below.

Definition 2. Deep Transfer Learning

For a transfer learning task defined by $\langle D_s, T_s, D_t, T_t, f_T(\bullet) \rangle$, if $f_T(\bullet)$ is a non-linear function which reflects DNN, then this is a deep transfer learning (DTL) task.

DTL is classified into four categories according to the technology used, namely instance-based DTL, mapping-based DTL, network-based DTL, and adversarial-based DTL, as shown in Figure 3.2 [7].

Sketch map of instances-based DTL [102], as shown in Figure 3.2 (1), assume that “although there are differences between the two domains, some instances in the source domain can be used by target domains with appropriate weights”. In

Figure 3.2 (1), the instances in light blue color in the source domain have dissimilar meanings to those in the target domain, and they are thus excluded from training dataset; the instances with dark blue color in the source domain have similar meanings to those in the target domain, and they are included in training dataset with appropriate weight.

Mapping-based DTL [103–105], as shown in Figure 3.2 (2), means to mapping the samples from the source and target domains into a new data space, where the samples from the two domains are similar and fit for a joint DNN; all samples in the new data space are utilized as a training set for the DNN. It is based on the assumption that “Although there are differences between the two origin domains, they can be more similar in an elaborate new data space.”

Network-based DTL [106–109], as shown in Figure 3.2 (3), assumes that a neural network is similar to the processing mechanism of a human brain, it is an iterative and continuous abstraction process. The front layer of the network can be regarded as a feature extractor, and the extracted features are versatile. First of all, the DNN uses a large number of data sets for training in the source domain. Subsequently, the partial network pre-trained in the source domain is transferred to be part of the DNN designed for the target domain. In the end, the transferred sub-DNN can be fine-tuned.

In Figure 3.2 (4), an adversarial-based DTL [110–114], inspired by the generative adversarial network, is introduced. It is based on the assumption that “for effective transfer, good representation should be discriminative for the main learning task and indiscriminate between source and target domain.” During training of massive datasets in the source domain, the front layer of the network is used for feature extraction. It extracts features from two domains and sends them to the adversarial layer, which distinguishes the origin of the feature. If the performance of the adversarial network is not good, it indicates that there is a slight difference between the two types of features and better transferability, and vice versa. In the following training process, the performance of the adversarial layer will be considered to force the DTL network to find general characteristics with high transferability.

In fact, the various techniques described above are often integrated to achieve better results. At present, supervised learning is more concentrated in recent research, but knowledge transfer through unsupervised/semi-supervised learning through DNN is

attracting more attention. Negative migration and measurement of transferability are important issues for traditional transfers. The impact of these two issues on DTL also requires further research. It is foreseeable that with the development of DTL, it will be extensively used for solving challenging issues.

3.1 TLCAF

Figure 3.3 shows the DTL fine tuning procedures, its output is a DTL transferred model. In this report, network-based DTL is used to explore building defect detection. as shown in the Figure 3.4. Here, faster R-CNN RPN network is applied for region of interest proposal, the image labeler as shown in Figure 2.3 for ground truth labeling, CNN network till fully connected layers are used for object feature extraction, and a linear classifier is used for prediction. The bounding boxes and prediction scores proposed by RPN are went through NMS for building defect detection and visualization.

3.2 A-TLCAF

Deep learning requires a lot of training data. However, only limited data are available for building defect assessment in this study; therefore, a DTL approach is explored. Feature-based transfer learning methods learn transformation by encoding

FIGURE 3.3: DTL fine tuning procedures

FIGURE 3.4: TLCAF

FIGURE 3.5: A-TLCAF

knowledge of a particular application. Advanced object features are particularly helpful for transfer learning tasks, if only limited labeled samples are available.

Architectural defects in building sector include cracks, finishing defects and hollowness, which have commonalities, such as low-level edges, textures and contour information. For specific features of higher layer advanced objects, DTL can be deployed for feature learning and defect assessment. In the following, an active transfer learning network based on convolution activation feature (A-TLCAF) is presented, as shown in Figure 3.5. Compared to TLCAF, the A-TLCAF system uses the top-N ROI proposals to automatically propose potential defects rather than image labeler. The image labeler is a tool that manually marks building defect category and its location in an image. “Active” in the A-TLCAF system refers to human actively labels top-N ROI proposals categories to augment our existing defect dataset. Compared with TLCAF, A-TLCAF can improve the verification

accuracy. The detailed process is carried out in four steps, as shown in the Figure 3.6.

- Faster R-CNN's RPN is used for region of interest proposal, and the ImageNet pre-trained ZFnet network is used for feature extraction. The source domain is the 20 category objects of Pascal VOC 2012. The target domain is a different type of architectural defect. In our experimental scenario, the source domain has no labels and the target domain has labels. The learned convolution activation feature was transferred from the unlabeled data (source domain, ImageNet's Pascal VOC 2012 dataset) using ImageNet pre-trained eight-layer ZF network.
- Features are extracted after fc7 with dropout layer to represent image-based architectural defects and non-defects.
- Training model for building defect detection and visualization
- Humans actively labels the categories of the top-N ROI proposals to augment our existing defect datasets and improve validation accuracy.

After training the model to represent the characteristics of image-based defects, a linear classifier (eg, SVM or softmax or random forest) is used for defect classification and prediction. Finally, a top N-level bounding box is drawn on the image at the position where the defect is detected.

Image output:
show bbox with top-N ranked score

Step 3

Defect prediction and visualization

Step 4

Actively learn top-N ranking features for fine-tuning

Step 2

Represent visual defect by learned higher-level features

Step 1

Transfer learning of convolutional activation features (TLCAF) from unlabeled data (Faster R-CNN pre-trained detection model)

Image Input

FIGURE 3.6: A-TLCAF network

3.3 Online-TLCAF

In real databases, the amount of data tends to increase progressively. Therefore, the learning method should be able to further train the model to extend the knowledge of existing models to incorporate the information from the new input data. The time cost for fine-tuning of a trained system is usually lower than the cost of retraining a system. It is not necessary to rebuild all knowledge bases when new data are added; instead, based on the original knowledge base, only the updates caused by new data are added; this method is called incremental learning, which applies to data streams and is more in line with human thoughts.

The aim of online incremental learning algorithms is to adapt to new data without forgetting the existing knowledge. Some incremental algorithms have some built-in parameters or assumptions that control the relevancy of old data, and learn representations of training data that are not even partially forgotten over time. Fuzzy ART [92, 94, 115, 116] and broad learning [99] are two typical examples. Broad learning was introduced, following the fast and dynamic learning capabilities of the random vector functional-link neural networks (RVFLNN) [98, 117–119]. It is different from ensemble learning. Ensemble learning could have many different models, but the data sources are the same. In contrast, broad learning fuses heterogeneous signals for synergistic knowledge discovery. Some special cases of broad learning include multi-view learning [120], multi-source learning [121, 122], multi-model learning [123], multi-domain learning [124, 125], etc. Transfer learning is to transfer information from one domain to another; if a domain is considered as a data source, then transfer learning is also a special case of broad learning. Learning and transfer are essentially inseparable, there will have knowledge transfer as long as a learning occurs, as any kind of learning must be influenced by the learner’s existing knowledge, skills, attitudes, etc. Transfer is the continuation and consolidation of learning, and it is the condition for improving and deepening learning.

There are many connection parameters in the last layer of FCN; hence, a time consuming retraining process is still required for DTL. Broad learning system (BLS) [99] helps boost the retraining process of the last layer, it is constructed in the form of a flat network, in which the original input is transmitted and placed as a “mapping feature” in the feature node, and the structure is widely extended in

FIGURE 3.7: Online-TLCAF for image-based defect detection

“enhancement”. An incremental learning algorithm was developed for rapid reconstruction for various extensions. If the network believes that expansion is required, then no retraining process is required. Incremental learning algorithm can be used for the increment of feature nodes and/or the increment of enhancement nodes; specifically, it allows incremental reconstruction of the system, rather than having to re-train the whole from scratch. The time scale singular value decomposition (SVD) approach can help simplify the final structure.

Supervised learning tasks (such as classification) often require good input representations to achieve superior performance. Feature representation not only represents data, but also captures the characteristics of the data. In general, pre-training model does not guide meaningful features for building defects, and TLCAF needs to be used to obtain linearly differentiated features. In the following, online-TLCAF is proposed.

- Figure 3.1 demonstrates the online-TLCAF workflow where the original image does not need to generate a true defect location marker. In Figure 3.8, the modified YOLO network is used for the objectness proposal, so that the objectness features are transferred to the new type of object.
- TLCAF for extraction of features of image-based building defects, as described in Section 3.2.
- Online-TLCAF model for image-based defect detection is demonstrated in Figure 3.7. TLCAF is used for feature extraction; broad learning and RVFLNN

are employed to eliminate the shortcomings of the long training/fine-tuning process, and provide generalization capabilities for function approximation; SVD is utilized to simplify the final structure.

In faster R-CNN framework, ROIs are manually labeled, for different anchors; IoU values are calculated according to comparison between anchors and ground truth boxes; and a linear classifier and NMS are employed to identify the optimal boundary for different objects. Labeling process for localizing new category objects, is very time-consuming, so is the computation of ROIs using the method in faster R-CNN. Therefore, $S \times S$ grid segmentation method of an image and objectness concept in YOLO [3, 4] is preferred, because this results in better detection performance than the ROI proposal method in the faster R-CNN (the speed could be more than 1,000 times faster, compared to some of the most popular methods [48, 49, 51]). For online TLCAF, it is proposed to use objectness as the first step for the proposal of region of interests.

In Figure 3.8 [3, 4], the image is divided into $S \times S$ units, where each unit predicts N bounding box, the confidence of objectness of these boxes, and the probability of C classes. Finally, to find the best predicted box, future adjustment is required to increase the confidence. Some units do not have any ground truth detection; in this case, the confidence of these boxes is decreased, while the class probabilities or coordinates are not adjusted. For each grid unit, it predicts four anchor boxes, and each box has a confidence score, and each grid unit predicts only one category, regardless of the number of boxes. Besides, YOLO's regional proposal

FIGURE 3.8: YOLO objectness proposal [3, 4]

FIGURE 3.9: Objectness proposals for image-based building defect

time-consuming bottleneck problem is solved in [3], and object detection accuracy is improved in [4]. Compared to faster R-CNN, transfer YOLO's objectness for ROI proposal is much faster and more efficient, due to less ROIs; another key merit of YOLO is that the cumbersome labeling work for object positioning is no longer needed, and the predictive knowledge of objectness can be completely transferred from source domain.

The objectness proposal algorithm is built based on YOLO v2, as shown in Figure 3.8, a threshold is set for $P(class_i) \times IoU(true | predicted)$ to determine whether it is an object. For image-based defect ROI proposal, as shown in Figure 3.9, the threshold is set to 0.001 by trial and error, and yellow bounding boxes predict the detected objects.

Broad learning is extended based on random vector functional-link neural network (RVFLNN). Compared to RVFLNN, broad learning does not directly use raw data as an input layer, it performs some conversion of the data, which is equivalent to feature extraction, and the changed features are used as the input layer of the original RVFLNN. It shows that features extracted by other models can be used to train a broad learning system, i.e. it can be used for machine learning algorithm assembly. Now we don't call broad learning system's first layer the input layer, but the feature layer.

The core of incremental learning is to use the last time calculation results, and the newly added data, update weights with a small amount of calculation. Broad learning uses the ridge regression algorithm for the first weight calculation, which

is slightly longer due to the iterative process. But the second, third... calculation involves only matrix multiplication, so the weight matrix is updated very quickly. Compared with the repetitive training of deep learning which often falls into the local optimum, the advantage of broad learning is obvious [99].

Online-TLCAF, as the embodiment of online deep transfer learning, supports the expansion of multiple feature nodes (filters in DNN), as well as enhanced nodes. Besides, incremental learning supports new incoming input data.

3.4 Model Evaluation

3.4.1 CIFAR-10

CIFAR-10 dataset is a set of images that are commonly used to train computer vision algorithms. It is one of the most popular datasets used in machine learning research. Since the images in CIFAR-10 are low resolution (32x32), this dataset allows researchers to quickly try different algorithms to see what works. Therefore, to evaluate the efficiency of online-TLCAF, CIFAR-10 dataset is used, four categories, including 'deer', 'dog', 'frog' and 'cat' are selected as shown in Figure 3.20. 1000 samples per class are chosen for training, and 300 samples per class are selected for testing. ResNet-50 is used for feature extraction through 'avg_pool' layer. The feature length is 2048 bytes. T-SNE is used to visualize four types of animal features in hyperplane. Figure 3.21 demonstrates the effect of TLCAF through T-SNE. It shows that the characteristics of the four animals are more aggregated and linearly separated.

To evaluate the performance of the proposed algorithms, VGG19-based TLCAF, ResNet-50 based TLCAF and online-TLCAF are applied to the CIFAR-10 dataset to compare the training duration and verify the accuracy.

VGG19 is a convolutional neural network that trains more than one million images in the ImageNet database. With a network depth of 19 layers, the image can be divided into 1000 object categories such as keyboards, mice, pencils and many animals. As a result, the network has learned a rich feature representation for various images. The image input size of the network is 224×224 . For VGG19-based TLCAF model training for CIFAR-10 selected dataset, set miniBatchSize to

FIGURE 3.10: VGG19-based TLCAF (CIFAR-10 dataset)

20, select SGDM for optimization, 'InitialLearnRate' is set to 1^{-4} , Max-epochs is set to 4000, and validation patience is 5. It stops because the validation patience reaches its threshold of 5. There are 5,000 training samples and 1500 test samples, and the training and verification process takes 16 minutes and 32 seconds. The test accuracy rate reached 91.33%. The training and evaluation process is shown in Figure 3.10. The chart shows that validation loss is similar to but slightly higher than training loss, there is no over-fitting issue. Figure 3.11 visualizes the characteristics of four animals trained using Google's deep Dream. Deep Dream is a feature visualization technology for deep learning. It visualizes the convolutional neural network (CNN). It synthesizes images that strongly activate network layers. By visualizing these images, the image features learned by the network are

FIGURE 3.11: Layer fc features (CIFAR-10 dataset)

FIGURE 3.12: ResNet-50 based TLCAF (CIFAR-10 dataset)

highlighted. It visualizes individual feature channels and their combinations to explore the pattern learned by the neural network. It allows the neural network to "amplify" the patterns in the image. These images are useful for understanding and diagnosing network behavior. The input is DTL trained model, the output is synthesized image.

Many visual recognition tasks benefit from deep CNN models. Therefore, there has been a trend to go deeper, solve more complex tasks and improve classification accuracy. However, as we go deeper, the training of CNN becomes difficult, the precision begins to saturate and then decreases. Residual learning tries to solve these two problems. The ResNet-50 is a 50-layer pre-trained residual network which has been trained on a subset of ImageNet database and won the ImageNet large scale visual recognition competition (ILSVRC) in 2015. In residual learning, it is not trying to learn some features, but trying to learn some residuals. The residual can be simply understood as the subtraction of features learned from the input of this layer. ResNet does this by using a fast connection (directly connecting the n th layer to the $(n + x)$ layer). It has been shown that training this form of network is easier than training a simple deep CNN, and also solves the problem

of reduced accuracy. During the ResNet-50 based TLCAF model training, SGDM was used for random optimization, miniBatchSize was set to 10, 'MaxEpochs' was set to 3000, 'InitialLearnRate' was set to 1^{-4} . Figure 3.12 shows its training and validation process. The training process took approximately 1 hour to complete. It achieved a test accuracy of 91.33%.

To evaluate online-TLCAF, ResNet-50 based TLCAF is used for feature extraction. For broad learning, there are four modes suitable for CIFAR-10 selected data sets.

For mode 1, one shot structure with fine-tuning under the back propagation (BP) algorithm, the l2 regularization parameter is set to 2^{-30} , the shrinkage scale of the enhancement node is 0.8, the feature node of each window is 10, the number of windows of feature nodes is 10, and the number of enhanced nodes is 200, run 10 epochs. The total training time is 0.748 seconds, the training accuracy is 94.12%, the total test time is 0.007 seconds, and the test accuracy is 90.13%. Table 3.1 lists the effects of the four modes of ResNet based online-TLCAF model.

For mode 2, one shot structure, the fine-tuning process under the BP algorithm is removed to save training time, the number of enhanced nodes is reduced from 200 to 50, and the epochs is reduced to 1 round. The total training time is 0.210 seconds, the training accuracy is 93.38%, the total test time is 0.049 seconds, and the test accuracy is 89.50%.

For mode 3, the increment of m enhancement nodes, the l2 regularization parameter and the shrinkage scale of the enhancement nodes do not change, the feature node of each window is 20, the number of windows of feature nodes is 10, and the number of enhancement nodes is 50, the number of enhancement node in each incremental learning is 10, the incremental learning step is 3, running for 1 period. The total training time is 0.208 seconds, the training accuracy is 93.48%, the total test time is 0.053 seconds, and the test accuracy is 89.70%. After two steps of incremental node, it takes 0.003 seconds to test, and the test accuracy is 89.60%. The steps of incremental learning and testing time are linearly positively correlated.

For mode 4, the increment of m2 + m3 enhancement node and m1 feature node, the regularization parameter for sparse regularization and the shrinkage parameter for enhancement nodes are unchanged, the feature node of each window is 10, and the number of windows of feature node is 10, the number of enhancement nodes is 50, run two epochs. "m1" indicates that the number of feature nodes for each

TABLE 3.1: ResNet-50 based Online-TLCAF results

	train acc(%)		test acc(%)		train time (s)		test time (s)	
	mean	std	mean	std	mean	std	mean	std
one shot structure with fine tuning under BP	94.12	0.11	90.13	0.25	0.748	0.135	0.007	0
one shot structure	93.38	0.09	89.5	0.11	0.21	0.004	0.049	0.002
increment of m enhancement nodes	93.48	0.03	89.7	0.35	0.208	0.002	0.053	0.001
- enhancement nodes in incremental step 2	93.59	0.12	89.6	0.17			0.003	0
increment of m2 +m3 enhancement nodes and m1 feature nodes	92.96	0.12	89.7	0.35	0.152	0.002	0.037	0.001
- enhancement nodes in incremental step 2	93.65	0.12	89.85	0.16			0.007	0.001

TABLE 3.2: Validation accuracy vs. training time for different ResNet architectures on CIFAR-10 [12]

Accuracy threshold	Time to threshold	Fastest model
0.918	1h 43m	ResNet20
0.93	4h 42m	ResNet56
0.94	10h 6m	ResNet164 (simple)
0.944	11h 42m	ResNet164 (bottleneck)

incremental step, it is set to 10; “m2” indicates the number of enhanced nodes associated with the incremental feature nodes for each incremental step, it is set to 15; “m3”, the number of enhanced nodes in each incremental learning, is 35. The step of incremental learning is 3. The total training time is 0.152 seconds, the training accuracy is 92.96%, the total test time is 0.037 seconds, and the test accuracy is 89.70%. After two steps of incremental nodes, the test takes 0.007 seconds to test and the test accuracy is 89.85%.

Table 3.2 is the benchmark to compare the validation accuracy vs. training time for different ResNet architectures on CIFAR-10 [12]. Fine tuning of the parameters can be traded off between accuracy and training time as needed. Table 3.3 indicating that online-TLCAF is the most effective way to learn new objects in real time while maintaining high discriminative power.

TABLE 3.3: Validation accuracy vs. training time for different TLCAF models on CIFAR-10

Accuracy threshold	Time to threshold	Best model
0.913	16m 32s	VGG19-based TLCAF
0.913	49m	Restnet50-based TLCAF
0.9	0.748s	Restnet50-based online-TLCAF

3.4.2 Building Defect Dataset

In the experiment, a database for image-based defects in a building is established, and the defects are divided into four major types, which are crack, corrosion, non-defect and finishing defect. For each type of defect, there are about 100-300 images. Finishing defect is a major drawback in the CONQUAS standard. According to BCA's recommendation, as a typical coating failure, finishing defect can be classified into discoloration, efflorescence, spalling and delamination, as shown in Figure 3.13, for future research.

FIGURE 3.13: Labeled building defects: crack, corrosion, non-defect and finishing defect

The data collection process is as follows.

1. Search and download some photos from Google Images and ImageNet
2. Some photos were collected from the construction sites, BCA CONQUAS rooms, our testing bed, campus, home, community and surrounding living environment.

The experimental procedure is listed below.

1. Build a dataset
 - 1). Collect data and build the building defect dataset.
 - 2). Learn field knowledge of different types of building defects from various literature reviews, books, websites and field experts.
 - 3). Build the dataset, crop and label the images.
2. Deep Transfer Learning (DTL) for building defect classification
 - 1). Use the 16 layers' vgg19 pre-training network (model A) as the base model, freeze the top 10 layers of the network, and change the softmax layer to meet the number of defect types.
 - 2). training network to obtain model B.
3. Using T-SNE for feature visualization
 - 1). Model A is used to extract features of the building defects.
 - 2). Visualize the features in the hyperplane.
 - 3). Use Model B to extract features of the same building defects.
 - 4). Visualize the features in the hyperplane.
4. Observe the effect of DTL in [Figure 3.14](#).
5. Visualize model B's features using Google's deep dream.

FIGURE 3.14: Deep transfer learning effect (our building defect dataset)

FIGURE 3.15: Comparison among different network-based TLCAF for defect validation accuracy

For different types of defects, the features extracted are visualized using t-Distributed Stochastic Neighbor Embedding (t-SNE). T-SNE is a well-known dimensionality reduction technique. It visualizes architectural defect capabilities. Figure 3.14(a) shows the defect feature extraction without TLCAF, Figure 3.14(b) visualizes defect feature extraction with TLCAF. The comparison between these two figures shows that the TLCAF can help to fine-tune the network and adapt the network weights to target domain (building defects), and generate more meaningful linearly distinguishable features.

Figure 3.15 compared the performance of different network-based TLCAF algorithms. It shows that ResNet is a good feature extraction tool - ResNet-based

algorithms present a good defect validation accuracy, which is larger than 0.8. Compared with VGG, ResNet has fewer parameters. The stochastic gradient descent algorithm can oscillate in the optimal direction along the steepest descent path. Adding a momentum item to a parameter update is one way to reduce this oscillation. Therefore, ResNet-50 and a stochastic gradient descent with momentum (SGDM) optimizer are chosen for TLCAF learning.

In Figure 3.6, faster R-CNN's RPN is used for TLCAF's ROI proposal. Faster R-CNN supports ZFNet and VGG-16. ZFNet was used for TLCAF and A-TLCAF experiments.

For online-TLCAF experiment. YOLO-v2 is used for object proposal, VGG, ResNet and Inception networks all can generate "TLCAF transferred model". Figure 3.4, Figure 3.5 and Figure 3.1 shows the diagrams and comparisons.

Since VGG is a serial network, and ResNet and Inception are directed acyclic graph (DAG) networks, VGG is supported for Google's deep dream feature visualization algorithm instead of ResNet and Inception network.

Figure 3.16 visualizes five types of learned features using Google Dream. It shows that the features for coating corrosion are some red dots on surface or edges, the features for crack are cracking curves, and those for discoloration features are blurred borders; the features for efflorescence are more prominent, compared with discoloration. For spalling and delamination, they appear to have longer crack lines connected together.

For five subdivided building defects, the model used is one of the best model trained through ResNet with randomly selected training data. It shows 81.3% validation accuracy rate. From Figure 3.13 and Figure 3.14 (c) and (d), some overlapping features were observed among discoloration, efflorescence, spalling and delamination. Hence, instead of increasing dataset, these three subdivided building defects are merged together, and we call it finishing defects.

The three categories of defects, including crack, corrosion and other defects, are then retrained by VGG19 and visualized in deep dream, the validation accuracy rate reaches 91%. Features of the three types of defects are visualized in Figure 3.16. It demonstrates the major differences between the three different defects. Figure 3.17 represents a confusion matrix for three architectural defects. We extracted the

FIGURE 3.16: Visualization of defect feature

		Accuracy: 90.83%			
Output Class	corrosion	93.5% 43	0.0% 0	2.2% 3	0.0% 0
	crack	0.0% 0	78.3% 47	5.8% 8	1.1% 1
	finishing defect	6.5% 3	20.0% 12	89.8% 123	0.0% 0
	non-defect	0.0% 0	1.7% 1	2.2% 3	98.9% 94
		corrosion	crack	finishing defect	non-defect
		Target Class			

FIGURE 3.17: ResNet-50 based TLCAF for building defects recognition

probability of each category and use one vs other to calculate the AUC, as shown in Figure 3.18, each AUC is greater than 95%. Due to the overlapping features between the crack and the finishing defect, the multi-class accuracy in Figure 3.17 is reduced to 91%. At present, self-taught weakly-supervised learning is applied to experiments. The involvement of field experts, data cleansing and enhanced data sets will further improve accuracy.

FIGURE 3.18: ROC of one vs others

According to [126], Inception-v4, also called Inception-ResNet-v2, which combines ResNet and Inception models, provides superior accuracy rate, regardless of the prediction time. Hence, Inception-ResNet-v2 was tested for feature extraction of image-based defects, the training process converges very slowly. Considered the trade-off between accuracy and training duration, ResNet-50 was selected for feature learning of image-based defects, its validation accuracy rate is 91%, same as VGG19, but ResNet-50 is more compact with less parameters, compared with VGG19 model. When training time is limited to 2 hours, Inception-ResNet-v2 optimization function can not converge within the time limit for three types of defects; therefore, five types of defects are not tested.

To adapt to different environments and changing world, it is expected to explore online learning algorithms for processing of high volume and high speed data. The comparison between the performance of online and offline models is shown in Table 3.4. VGG19 and ResNet-50 both take about half an hour to finish the training; in contrast, online-TLCAF only takes 0.3 second to learn the same amount of features. Online-TLCAF also acquires the highest accuracy rate for recognition of both the 3 types and 5 types of defects.

Input modes were selected for incremental learning by batch, each with 200 functions. The number of input modes added per incremental step is 200, and the

FIGURE 3.19: Comparison of complexity for different CNN models [8]

number of enhanced nodes added per incremental step is set to 10. It takes four steps to complete the incremental learning of the training. The regularization parameter for sparse regularization is set to 2^{-30} . Other parameters utilized include: 0.8 for shrinkage parameter for enhancement nodes, 10 for the feature nodes of each window, 10 for the number of windows of feature node, and 50 for the number of enhanced nodes. According to the current data set, the test accuracy for the three types of defects reached 95% with **0.3-second training time**.

Experimental results of the online-TLCAF on the first batch building defect dataset including crack and finishing defects are compared with other network-based TLCAF models in Figure 3.15. For the validation accuracy rate in the figure, the learning rate is set to 0.0001. Figure 3.19(a) shows the top-1 accuracy of the most relevant entries submitted to the ImageNet Challenge, from AlexNet [65], to the best performing Inception-v4 [127]. The latest ResNet and Inception networks provide better validation accuracy rate. Figure 3.19(b) gives a different but more informative view of accuracy rate, it also visualizes computational cost and the number of network parameters. It shows that VGG is by far the most expensive architecture - both in terms of computational requirements or the number of parameters[8]. Most of the existing models need around 10 minutes to several hours to converge, and ResNet-50 + random forest (RF) (60 nodes) is relatively fast for training of a new model for small sample object recognition. It only takes

TABLE 3.4: Comparison between online and offline models

	VGG19	ResNet-50	online-TLCAF
3 type defect	91%	91%	95%
5 type defect	76.74%	77.33%	79.3%

FIGURE 3.20: CIFAR-10 selected dataset

FIGURE 3.21: Visualization of TLCAF trained CIFAR-10 features through T-SNE

about 3 seconds for training; however, it does not support incremental learning. In contrast, the proposed online-TLCAF network, which is built up based on TLCAF through ResNet model plus broad learning (BL), yields a much better validation accuracy rate than all of the existing models/methods. Moreover, it only takes about 0.2 seconds for retraining process, and it supports deep structures online incremental learning for interaction with environment. Hence, the online-TLCAF network demonstrates the effectiveness of online DTL.

Table 3.5 shows that there are six modes for online-TLCAF as shown in Table II. From the results, it is observed: According to the current data set, the test accuracy for the three types of defects reached 91% with 0.3-second training time. From Table 3.6,3.7,3.8,3.9, it is observed that incremental of m input patterns does not help on testing accuracy because the features are used directly as input rather than raw data.

TABLE 3.5: Different modes for online-TLCAF

mode	description
1	one shot structure with fine tuning under BP
2	increment of m input patterns
3	increment of m enhancement nodes
4	increment of m2+m3 enhancement nodes and m1 feature nodes
5	increment of m input patterns and m2 enhancement nodes
6	one shot structure

TABLE 3.6: Pre-trained VGG19-based online-TLCAF

mode	train acc (%)	test acc (%)	train time (s)	test time (s)
1	93.08	84.62	0.443	0.003
2	70.86	48.92	0.064	0.058
3	90.63	86.00	0.269	0.084
4	91.56	84.42	0.179	0.072
5	70.38	50.89	0.140	0.088
6	89.54	86.19	0.266	0.077

TABLE 3.7: Trained VGG19-based online-TLCAF

mode	train acc (%)	test acc (%)	train time (s)	test time (s)
1	100.00	91.72	0.520	0.003
2	89.00	56.02	0.060	0.056
3	99.92	91.52	0.259	0.081
4	99.83	91.32	0.178	0.070
5	88.38	58.38	0.142	0.085
6	99.75	91.52	0.258	0.076

Table 3.6 and Table 3.7 compared pre-trained and trained VGG-based online-TLCAF, the trained VGG19-based online-TLCAF improves validation accuracy by more than 5%. The TLCAF network is helpful.

Tables 3.8 and Tables 3.9 show approximately 1-2% improvement before and after TLCAF of Resnet-50. In addition, comparing Table 3.8 and Table 3.6, we can see that the pre-trained Resnet-50 online-TLCAF's test accuracy is 5% higher than pre-trained VGG19 online-TLCAF. It shows that the ResNet-50 can capture more meaningful generic features than VGG19.

Online-TLCAF acquires the competitive accuracy for three types of building defects. Since the features extracted from ResNet-50 is 2048 bytes instead of 4096

TABLE 3.8: Pre-trained Resnet-50-based online-TLCAF

mode	train acc (%)	test acc (%)	train time (s)	test time (s)
1	95.64	89.94	0.366	0.002
2	73.57	49.41	0.037	0.029
3	93.21	87.87	0.183	0.046
4	94.53	88.46	0.117	0.039
5	77.13	50.59	0.085	0.052
6	93.57	89.35	0.178	0.040

TABLE 3.9: Trained Resnet-50-based online-TLCAF

mode	train acc (%)	test acc (%)	train time (s)	test time (s)
1	97.34	90.83	0.327	0.002
2	78.29	49.70	0.038	0.029
3	95.42	90.83	0.169	0.046
4	96.16	90.83	0.116	0.038
5	77.88	51.18	0.082	0.049
6	94.76	90.53	0.178	0.041

bytes in VGG19, ResNet-50-based online-TLCAF training is almost twice as fast as VGG19-based online-TLCAF. It shows that online-TLCAF is the most effective way to learn new objects in real time while maintaining high discriminative power.

3.5 Summary

This chapter runs through the entire image-based object detection process. This starts from data collection, feature extraction, followed by feature learning and visualization, as well as data cleansing and data augmentation. The A-TLCAF network and online-TLCAF network were proposed as new machine learning methods, especially for small sample new category object recognition. At the same time, the objectness of the ROI proposal is integrated with A-TLCAF to save cumbersome and inefficient pre-labeling work, and IoU related linear regression algorithms used for object localization. Moreover, the online-TLCAF network is developed for incremental learning of high-dimensional dynamic data streams.

The proposed method, Figure 3.7, adds new factors to the traditional TLCAF. Figure 3.4, Figure 3.5 and Figure 3.1 are used to illustrate the relationship and differences between traditional TLCAF, A-TLCAF and online TLCAF.

Compared to TLCAF, online TLCAF has evolved with the following two functions:

- The ROI proposal is being replaced by an automatic objectness proposal instead of manual image labelling.
- Linear classifier in traditional TLCAF is replaced by extensive online incremental ones.

Compared to TLCAF, A-TLCAF auto proposes Top-N ranking ROI proposals for human active labelling to improve validation accuracy.

In order to process data at high speed and quickly adapt to data changes, an online TLCAF is proposed. A-TLCAF and online TLCAF can complement each other, to enable active, continuous and efficient learning through interaction with the real environment.

Chapter 4

Technical Contribution for Defect Detection

Defect inspection is an important part of modern industry, there are two disadvantages for traditional methods: 1) their performance depends on the design quality of the handcrafted features, but it is difficult to design these features in advance; 2) if the training and test data are from the same distribution, they work well, but this hypothesis is not fulfilled in many industry applications. DTL was developed for image-based defect detection, because deep learning is able to extract the hierarchical representation of raw data; at the same time, transfer learning is a good way to learn different but related tasks.

In this chapter, a robot system for post-construction quality assessment is proposed. to speed up inspections, and auto-generate more reliable inspection reports. It targets at fully automated inspections to address crack, hollow and finishing defects, and it is possible to reduce labor costs and achieve higher accuracy.

Significant progress has been made in image-based object detection, primarily due to the availability of large datasets, the revival of deep learning, and the rapid development of computing hardware. Deep CNNs have been able to learn data-driven representative image features of massive training samples; however, it remains a challenge to access large data sets in the field of building quality assessment. For the proposed TLCAF network for image-based defect recognition, it uses RPN for ROI proposal, deep CNN for feature extraction, linear classifier for feature prediction, and active learning of Top-N level ROI for fine tuning. Massive validation

experiments are performed in the CONQUAS room and construction site, and the novel automatic evaluation method proposed shows satisfactory performance in crack detection, hollow and finishing defect detection. To the best of the author's knowledge, this is the first attempt to develop a visual inspection system for building quality assessment.

This investigation is of interests not only to researchers working on general-purpose automatized applications, but also to those on automated building quality inspection. The knowledge and insight acquired may encourage future research in this area.

4.1 Crack Detection - Work in NTU

Manual inspection by human eye for inspection of defects in a building, is subjective and relies on the experience and spiritual focus of the inspector. The development of automatic defect assessment system provides a solution to these shortcomings. In [128], the crack segmentation parameters were adaptively adjusted based on the depth parameters, the crack features were extracted by linear discriminant analysis (LDA); the performance of SVM, nearest neighbor classifier and neuron networks (NN) were compared for crack recognition. An automatic crack quantification method was developed in the study. In [129], pyramid SVM is employed to detect cracks in cookies after Wilk's λ analysis, it shows the effects of hough-based features on crack recognition. In a method developed for bridge deck inspection [130], Laplace Gauss algorithm was used to detect cracks. Automated decision making for defect identification can be performed by establishing signal thresholds or defining defect features. ROIs on the object can be extracted to provide a characterization of defect. The features extracted need to be trained for improvement. NN can be used to train crack features, including subtle changes; special training on the network is required for each type of defects.

4.1.1 Approach

In the system proposed, faster R-CNN is applied. It integrates RPN and fast R-CNN detector. The RPN reduces the proposal for the top N proposed areas, via

the use of objectness score as a “attention” mechanism for the unified network; subsequently, the RPN outputs ROIs to fast R-CNN for defect prediction. For input images, the RPN generates a group of ROI proposals, each proposal has an objectness score [55].

The ROI proposals are parameterized relative to the bounding boxes (bboxes). Objectness is often used to quantify the probability of a group of object classes [51, 70]. An eight layer Zeiler and Fergus network (ZFNet)[131] is employed in the study. The RPN module has two sibling output layers, one for the discrete object probability prediction (per region of interest) and the other for the bounding box regression offset.

Neurons in different CNN layers act as local filters. The rectifying linear unit (ReLU) is employed as an activation function for each CNN layer instead of sigmoid function, since ReLU can solve the gradient disappearing issues. When using neurons with infinite activation such as ReLU, the normalized layer of local response (ie, local contrast normalization) proves to be useful. It stimulates competition of large-scale activities in nearby neuron populations. ReLU reduces the dimension of hyperparameters and shows good performance in suppressing overfitting. For subsampling, max-pooling is used, which has a better accuracy in finding object outlines, compared to average-pooling. CNNs are trained by backpropagation and SGD. The convolution proposal layer is shared by RPN and fast R-CNN. The other two sibling layers are the object scoring layer and the bounding box regression layer of the proposal. ROIs determined by RPN are passed to fast R-CNN for feature extraction, 4096 byte features are extracted after two more fc layers with dropout mechanism used to improve generalization. Dropout is a technology that prevents NN over-fitting by preventing complex synergistic adaptation of training data. After passed through fc7, each proposed 4096-byte feature is sent to a linear SVM for classification. The positive predictions are then integrated with their bounding boxes for defect reconstruction and visualization. Moreover, [131] shows both CNN layers and fc layers both help in reducing classification errors. This is why the features are extracted through fc7 layer in this study. At the same time, the increase in the filter in the convolutional layers 3, 4, 5 also improves the classification accuracy [131]. In addition, data augmentation and further fine-tuning are also performed to increase the classification accuracy.

FIGURE 4.1: A-TLCAF learning

The TLCAF training steps are as follows: faster R-CNN with pre-trained model → feature extraction (4096-byte feature is extracted after fc7) → linear classifier → Well-trained model.

The testing steps follow: faster R-CNN pre-trained model → extraction of features with ROIs → trained linear classifier for prediction → detection.

In order to improve the recognition rate, fine-tuning is expected to actively learn the Top-N ranking ROIs. A larger “N” helps locate image-based defects with a more accurate box. Typically, “N = 3” is enough for the machine to learn the defect features. In this work, “N = 20”, was used to extract useful features of image-based defects. Figure 4.1(a) gives an instance of the testing results before fine-tuning, and Figure 4.1(b) demonstrates the results after active learning. It shows the impact of active learning on defect detection.

4.1.2 Feature Visualization

In [131], feature visualization shows that the feature maps of lower CNN layers represent edges and countours, and the feature maps of higher CNN layers represent a combination of high-level patterns. Cracks and finishing defects result in abrupt gradients in edges, colors and textures; most of the meaningful features are low level CNN features. Figure 4.2(a) demonstrates a feature map of the first two CNN layers for damage and jointing defects. It shows that CNN’s shallow network can extract and learn meaningful edge and shape defect features. Figure 4.2(b) shows some typical defects with the strongest feature mapping. In the study, horizontally flipped images were used for data augmentation. It shows that the

FIGURE 4.2: Feature map of different defects

default pertained features within faster R-CNN is able to extract low level defect feature information with high signal-to-noise ratio.

4.1.3 Defect Prediction and Detection

Color camera is used to capture crack and finishing defects. Use the method proposed in Section 3, a visualization of image-based building defects through principal component analysis (PCA), is presented in Figure 4.3. The PCA performs dimensionality reduction by projecting the entire training to a subspace that maximizes data information. To assess the separability of this class, PCA was used to visualize the first three principal components (PC) of the 2000 random observations. In Figure 4.3, the positive features are displayed as yellow dots in the 3D space, negative features are represented by the blue dots. It shows that images of different categories are linearly separated in the hyperplane.

In this dataset, 1041 positive features were extracted from 680 image, which contained cracks and finishing defects; 80% of the 1041 features were randomly selected for the training set, and the remaining 20%, i.e. 201 features, were put into the test set. For non-positive features, 49781 features were extracted from 610 indoor images; 80% of the features are randomly put into the training dataset, with the remaining 20 % of the features, i.e. 9922 features, are put into the testing dataset. 25 tests were performed. During the experiment, balanced F1 measurement (F_score)

FIGURE 4.3: Visualization of image-based defect features

(ranging from 0 - 100%) was used to represent the harmonic mean of precision and recall, higher F_score shows better performance. The best model reaches 99.99% accuracy, with F_score of 99.76%. The average accuracy of the 25 trained TLCAF models is $99.62\% \pm 0.33\%$, and the average F_score is $91.64\% \pm 6.49\%$.

4.1.4 Experimental Results

Figure 4.4 demonstrates the defect detection results after integrating RPN and TLCAF. Table 4.1 shows the TLCAF result for crack validation result using faster R-CNN. For testing set of different building defects, A-TLCAF is a complementary approach to improving TLCAF. A-TLCAF using RPN of faster R-CNN to automatically propose Top-N region of interest of different defects.

FIGURE 4.4: Detection result [9]

TABLE 4.1: Comparison of crack identification algorithms

crack classification		sensitivity(%)	specificity (%)	DP
TLCAF	best model	99.51	100	+infinity
	mean	96.79	99.68	2.19
λ analysis + SVM [129]		96	98	1.69
crack prediction [128]	LDA + NN	84.1	74.5	0.66
	LDA + SVM	84.1	72	0.63

TABLE 4.2: Crack detection accuracy

models	ROI	crack detection accuracy
TLCAF	pre-trained RPN	93.15%
A-TLCAF	pre-trained RPN to propose	100%
	Top-N ROI for fine-tuning	

Totally 53 positive images and 166 non-positive images were tested, 93.15% accuracy and 98.09% precision were achieved. A balanced 80.65% precision is acceptable although more false positive images are reported, the loss of important true positive images is avoided. The prediction threshold of the classifier can be adjusted based on assessor's requirement. These are the results before active learning. With active learning of the Top-N ranking ROIs, the accuracy can be improved significantly. As shown in Table 4.2, fifteen images which cannot be identified by TLCAF, all are well recognized by A-TLCAF.

Compared with some traditional algorithms, including adaptive crack detection [128] and Wilks λ analysis + Pyramid SVM [129]. The TLCAF algorithm developed in this study shows a better performance in crack detection.

In the field of machine learning, discriminant power (DP) is usually applied to evaluate algorithm's performance in distinguishing positive samples from negative ones. For an algorithm, $DP < 1$ means that it is a very bad discriminator, $DP < 2$ indicates a limited one, and $DP < 3$ means a fair one; in other cases, the algorithm is a good discriminator [132].

$$DP = \frac{\sqrt{3}}{\pi} \times (\log X + \log Y) \quad (4.1)$$

$$X = \frac{\textit{sensitivity}}{1 - \textit{sensitivity}}; Y = \frac{\textit{specificity}}{1 - \textit{specificity}} \quad (4.2)$$

Table 4.1 shows that TLCAF network produces the best discriminant power, indicating that it is the best discriminator among the crack detection methods evaluated. The DP value indicates good performance of TLCAF - the randomly trained TLCAF model is a fair discriminator and the best TLCAF model is a good one. In contrast, Wilk's λ analysis + pyramid SVM method produces a limited DP , the adaptive crack detection method using LDA + NN/SVM, only provides poor DP s.

The traditional methods mentioned in Table 4.1 can only detect cracks, but architectural defects include not only cracks, but also corrosion and different types of finishing defects. Compared with traditional methods, CNN features can extract and learn 2D shape features, texture features and color features of different types of image-based defects. However, the limitation is that DNN methods require a large amount of data. Due to limited data samples, especially hidden defects captured by thermal imaging cameras are even more less, a TLCAF model is proposed to learn defect characteristics from a small sample data.

Meanwhile, in real life, the amount of data tends to increase gradually. The purpose of the online incremental learning algorithm is to adapt to new data without abandoning existing knowledge in a faster way. Therefore, an online TLCAF model for autonomous building defect assessment is proposed.

4.2 Hollowness Assessment - Work in NTU

Among available NDT assessment techniques, Infrared thermography (IRT) is widely used in building diagnosis. Compared to other NDT methods, it is safe, fast, and suitable for inspecting targets with a large area. A thermal camera can

be mounted on a moving vehicle at a speed of 20 km/h to inspect an irregular terrain [133]. As a most effective NDT technology, active IRT robotic system has been used for defect inspection for large and complex composite structures. IRT is much better than traditional sampling method using a hammer, especially for evaluation speed.

Regarding the thermal diagnosis of defects, region of interest is selected by pre-trained feature descriptions. In the literature, the easiest way to distinguish hotspots/coldspots on thermal images of buildings is to use morphological image processing techniques and statistical methods to identify the differences between hotspots/coldspots and reference temperatures [134, 135]. Various intelligent techniques, i.e. artificial neural network (ANN) [136, 137], SVM [138, 139] and neuro-fuzzy algorithm [140], have been applied for classification tasks. For instance, in [141], the author used ANN classification to extract crack patterns from three different images, i.e. grayscale, color, and thermal image. The results show that IRT image provides the highest classification accuracy, compared to grayscale and color images.

Most IR images are processed using pixel-based functions, the algorithms are evaluated through visual perception and defect prediction [142–145]. In existing algorithms, wavelet transform allows the process to be located in the time domain, but the potential of the technology has not been fully investigated. SVD and PCA can be used to reduce dimension, they neglect low-order components, but their defect features are non-trivial [146]. Hough/radon transform distinguishes defects from non-defects by assuming temperature linear behavior in log-log coordinates; it does not require a reference point, but can only be applied to a semi-infinite sample. Histogram of oriented gradients (HOG) and Scale-invariant feature transform (SIFT) algorithms calculate gradient amplitude to show points of interest, with respect to the direction or dimension. Extraction of the points of interest provides a “feature description” of the defect. These features require training for improvement, but the feature itself does not have self-learning ability. Neural network is able to learn crack characteristics, including subtle changes; however, each type of defect requires special training on the network. Faster R-CNN [55] fits for near real-time object detection; similar to neural network, special training is required.

4.2.1 Approach

IRT is reliable for inspecting the quality of the building surface, because it is very sensitive to changes in surface temperature. In this method, compared to intact surface/subsurface, discontinuous temperature distribution appears at the location where defects exist. A comparison on studies using IRT inspection is presented in Table 4.3.

Material	Concrete, carbon fiber and glass FRP composites [147]
Method	PT; four 0.5 kW halogen heaters, 60 second square heat pulse; FLIR PM695 camera, long wave, with sensitivity of $<0.08^{\circ}\text{C}$ at 30°C .
Findings	An effective method to minimize uneven heating effects, and evaluate the relative depth of FRP/concrete interface manufacturing defects.
Material	Marble [148]
Method	PT; 1.5kw halogen lamp, heating (20 minutes) and cooling (30 minutes) transients at 20 second intervals; FLIR T640 camera, long wave, with sensitivity of $<0.035^{\circ}\text{C}$ at 30°C
Findings	Application of digital codes in PT results in higher precision for the anomalous properties and geometry of building structures.
Material	Concrete, marble, ceramic [149]
Method	PT; recording time: 116 min; external stimulus: PT, the oven is heated at 50°C for 2 hours; recording intervals (s): 60 s; FLIR PM695 camera, long wave, with sensitivity of $<0.08^{\circ}\text{C}$
Findings	Sensible heat release was verified. The monitoring method is consistent with traditional energy equation.
Material	Concrete, FRP composites [150]
Method	PT; 1kw heater; apply flexible electric cover for 50 seconds; IR camera
Findings	IRT can detect adhesion defects.
Material	Concrete structures [151]
Method	PT and LT; heat samples in an oven at 90°C for 3 hours; FLIR T360 camera, long wave, sensitivity $<0.06^{\circ}\text{C}$ at 30°C
Findings	Vertical cracks usually have a very thin shape, and IR cameras with high thermal sensitivity are required to detect them.
Material	Mosaic and ceramic tile [152]
Method	Long pulse exposure IRT (LPT); 2 kw thermal radiation, exposure of 130 min; IR camera
Findings	The second derivative calculation provides information on hollowness depth.
Material	Concrete samples [152]
Method	LPT; 2 kW thermal radiation, exposure of 130 min; IR camera
Findings	After applying a transient pulse excitation, several surface cracks appear as black lines in thermal image.
Material	Ceramic tile [153]

Method	TPT; 2.4kw infrared radiators, 3-min heating; IR camera
Findings	Successful detection of hollowness beneath the tiles.
Material	Carbon/epoxy and glass/epoxy composites [154]
Method	1 kW round quartz lamp; infrared camera, with sensitivity of $<0.07^{\circ}\text{C}$
Findings	LT is better than PT. The contrast decreases if the size of the defect decreases and the depth increases.
Material	Wood [155]
Method	Microwave heating for 2 min; IR camera
Findings	Thermal models presented with encouraging experimental results.
Material	Building walls made of brick and cement plaster [156]
Method	solar irradiation; IR camera
Findings	Although the processing time is increased, the defect evaluation in the transient state is superior to the steady state.
Material	Non-reflective ceramic tiles (ours) [9]
Method	LT; 3kW quartz heater, heating of 20 seconds; sliding window method for uniform heating; FLIR A310 camera, long wave, sensitivity of $<0.05^{\circ}\text{C}$
Findings	Successful detection of hollowness under tiles, in terms of shape and size.
Material	Reflective ceramic tiles (ours) [9]
Method	LPT; 3kW quartz heater, heating of 3 minutes; FLIR A310 camera, long wave, sensitivity of $<0.05^{\circ}\text{C}$
Findings	Successful detection of hollowness.
Material	Marble, granite tiles (ours) [9]
Method	LPT; 3kW quartz heater, heating of 5 minutes; FLIR A310 camera, long wave, sensitivity of $<0.05^{\circ}\text{C}$
Findings	Successful detection of hollowness under tiles, including its shape and size.
Material	Structure defect (ours) [20]
Method	LPT; 1.5kW halogen heater, heating of 20 minutes; FLIR a655sc camera, long wave, sensitivity of $<0.03^{\circ}\text{C}$
Findings	Successful detection of concrete voids of different dimensions and under different depths of concrete cover.
Material	Hidden coating corrosion (corroded steel plate with epoxy coating) (ours) [21]
Method	LPT; 1.2kW hair dryer, heating of 3 seconds; FLIR a6702sc camera, mid wave, sensitivity of $<0.02^{\circ}\text{C}$
Findings	Successful detection of the hidden corrosion under the coating, including shape and size.

TABLE 4.3: Hollowness detection methods review

FIGURE 4.5: Hollowness feature map

4.2.2 Experimental Results

A preliminary study was conducted using a FLIR A310 camera to explore the performance of thermal imaging cameras in hollow detection. The indoor thermal image captured (Figure 4.5(a)) shows the void under the tile. The tested ceramic tiles (no reflection) were heated by a periodically energized locked thermal radiation heating method, using a 3 kW quartz heater for 20 seconds. Due to the low thermal conductivity of ceramic materials, long pulse exposure can cause uneven heating, result in misunderstanding of the results. Therefore, Heater's sliding window movement is preferred to reduce the effects of non-uniform heating. In the thermography, lighter cells stand for the presence of voids under the cells since they are less dense, so they heat faster than solid-filled and dense cells. This indicates that the thermal response is effective for hollow detection. Figure 4.5(c) shows voids and joint defects detected under the granite floor of the CONQUAS room after 5 minutes of continuous heating of the tiles. Figure 4.5(b) and (d) visualize a stronger feature map in the second CNN layer of the faster R-CNN.

To evaluate the limitations of thermal method on illumination and reflection, ceramic walls with reflection were tested. The hollow can be seen in the thermal

FIGURE 4.6: Thermal images of hollowness

image before heating; after heating the tiles, measurable size and shape of hollow can be obtained. Figure 4.6 presents experimental results for reflective ceramic walls and granite floors. Figure 4.6(a) demonstrates the thermal image captured at room temperature (without heating) - a medium-sized circular hollow and a large elongated hollow (diameter ≥ 8 cm) can be seen, but the third small hollow with a diameter of ≤ 5 cm is not visible. Figure 4.6(b) is the results after 3 minutes' heating. A medium-sized circular hollow and a large elliptical cavity can be clearly seen, the size matches with ground truth dimension, the third small hole can also be seen, although not as clear as the other two bigger hollownesses. Due to the small size, the third small hollow is considered to be unimportant. These figures show that the proposed method is promising for hollowness detection.

Another test was undertaken on the granite floor with hollowness, which is built in the testbed. Figure 4.6 (c) and (d) shows that after a continuously heating of the floor for 3 minutes, using a 3kW quartz heater, the holes and boundary defects

FIGURE 4.7: Hollowness detected

hidden behind the surface were clearly obtained by a long-wave thermal camera (FLIR A310). Figure 4.6(c) presents the hollow and boundary defects identified below the surface, and Figure 4.6(d) shows the different sizes of hollows detected under the granite floor. These results show that IRT is an effective method in detecting hollows of various sizes and shapes. Using the A-TLCAF network, hollow features in thermal images can be correctly detected by machine learning algorithms in the system. Figure 4.7 shows a typical example for hollowness detected.

4.3 A-CONQUARS Robot System

4.3.1 Surface Measurement

3D scanners are highly accurate scientific devices that could be used for surface measurement, they are classified based on their working principles, i.e. contact and non-contact scanners. The former is accurate and widely preferable in industry. However, as the scanner must be in contact with the object being scanned, there is a possibility to cause damage to the object. For non-contact scanners, they are further classified into active and passive scanners. A passive scanner does not emit any radiation itself, but measures the radiation reflected by the surface of an object to achieve the geometry. The majority of passive scanners are based on visible light. In active scanning, extra energy (e.g. visible light, high-energy laser beam, ultrasound and X-rays) is projected onto the object and the reflected energy is captured to calculate the three-dimensional spatial information. Active

FIGURE 4.8: Perpendicularity measurement

FIGURE 4.9: Flatness measurement

3D scanners that use structured lighting or modulated lighting for measurements, have high accuracy, but they are too expensive, compared to 2D laser scanners.

In the experiments done by Trimble Navigation, Ltd. in the lab of the authors, measurements of surface flatness and perpendicularity between two selected surfaces were performed using a Trimble X130 3D laser scanner and Trimble Real Works application. Figure 4.8 shows the measured angle between the two selected surfaces, and Figure 4.9 shows the portion of the wall (the area in blue color) where the flatness exceeds the allowed tolerance. Our Quicabot v4 has been able to complete flatness measurements in real time [157].

3D laser scanners have very high accuracy, but they are much more expensive than 2D laser scanners. 3D thermal scanner is cheap and able to generate real time 3D point cloud data; however, its accuracy is relatively low and affected by environment[158]. 3D structure light scanner (for example, kinect v2 for indoor

measurement) is cheap, fast, accurate, and eye safe. The invisible structure light (ISL) which it uses, permits one-shot reconstruction, and detection of surface and shape defects ; however, the scanner is sensitive to ambient light, and cannot be used to scan shiny surfaces [159]. The traditional 2D laser scanner is cheap, robust, and not affected by ambient light. It can scan any material, color or luster, and provides excellent accuracy. The disadvantage is that it can only provide 2D data. For laser-based scanners, the laser beam brings safety issue and precautions must be taken to prevent contact of laser beam with eyes of user [160]. Hence, an eye-safe 2D laser scanner, together with an electronic inclinometer, is recommended for alignment and evenness measurement.

4.3.2 Sensor Fusion and Integration

Mobile robots are mature for construction quality assessment in an autonomous manner. The challenge of robot-based quality assessment is autonomous safe navigation in an unknown environment. Nevertheless, it is practicable to build a new robotic system to automate quality assessment. through use of appropriate instruments, together with further developments in robotic technology.

Accurate environmental representation is crucial for navigation. The first type of navigation method is global path planning on a known environment that requires the development of a predicted path from the start point to the destination before the robot begins to move. On the other hand, local path planning is defined in an unknown environment, and built-in algorithms are able to identify the path of the robot from the origin to the target point. A review of mobile robot navigation technology research is presented as follows.

Traditionally, the route map method is used to connect the starting point to the destination, the generated path acts as a road network for the mobile robot [161, 162]. Unlike route map approach, cell decomposition methods distinguish between obstacles and geometric free zones, support the planning of robot teams in 2D regions [163, 164]. In machine learning field, the target and obstacle positions are designated as attractive and repulsive forces, respectively. It causes the robot to move towards the target and be “pushed open” by the obstacle. The issue is local minima may cause errors [165, 166].

For heuristic methods, fuzzy logic theory was deployed for behavioral design and coordination. Fuzzy logic controllers are used to benefit from human expertise and the ability to handle sensor data errors [167]. As an ANN method, the intelligent four-layer NN can better avoid obstacles and solve the navigation problem of mobile robots. [168]. Firstly, the adaptive neuron fuzzy technique generates automatic navigation in an unstructured environment, wherein the speed of the mobile robot is adjusted following the distance information processed by the neuron fuzzy controller [169–171]; secondly, genetic algorithms use quadtree data structures to create environmental databases and generate optimal paths from these databases; it avoids complex training of ANNs and local minimum errors [172, 173]. Particle Swarm Optimization is designed to recognize the optimal path, it adapts navigation to the environment [174–176]. A comparison among various robot navigation techniques was presented in [177], it indicates that heuristic navigation methods are more efficient for unknown and dynamic environment, compared to traditional model-based methods.

Multi-sensor fusion and integration (MFI) is an important portion of robot systems. The advantages and limitations of the most commonly used MFI methods and their integration are discussed below.

Kalman filter (KF) uses a series of measurements that contain noise observed over time, which produces a more accurate state estimate than a state estimate based on a single measurement. Due to the recursive nature of the algorithm, it can be run in real time using only the current input measurements and the previously calculated state and its uncertainty matrix; no additional past information is needed [178]. The study in [179] uses extended and unscented KF (EKF and UKF) to fuse odometer data and sonar sensor data to compensate for the cumulative error of the odometer and the inaccuracy of the ultrasonic sensor. Their work shows that ranging calibration and sensor fusion can significantly improve the local positioning capabilities of the robot.

The classification method in MFI groups multi-source data into classified data sets, using parameter or non-parametric methods. Compared to parametric methods, nonparametric classification methods are not limited by prior assumptions about the distribution of input data. In the classification method, SVM is often used in mobile robots for simultaneous location and mapping, object detection, navigation

and anti-collisions. SVM is suitable for secure navigation, because it uses the maximum margin concept [180, 181].

For inference methods, Dempster-Shafer (D-S) theory and bayesian inference are two prevalent advanced inference algorithms for MFI. Bayesian inference can effectively solve most fusion problems. It serves as an abstraction to provide a probabilistic framework for recursive state estimation. In [182], it is proved that the D-S reasoning theory can solve some issues which cannot be solved by probability theory. Particle filter (PF) can represent any probability density in a nonlinear dynamic environment because it concentrates on regions with high probability. D-S theory can simulate beliefs about basic assumptions and compound hypotheses, and often uses other algorithms to improve decision accuracy [183, 184].

MFI promotes high quality building quality diagnostics for the A-CONQUARS system, including positioning and navigation. Related context-aware tasks include obstacle avoidance and defect detection by merging the vision sensor with the range sensor. Lidar is used to provide distance information, and visual sensors recognize objects to track and sense the relative position between the robot and the environment[178]. At the same time, inertial sensors and odometers are used for positioning and navigation. In [172], genetic algorithms are implemented for sensor data fusion for efficient navigation. From the literature review, most fusion algorithms are limited by the computational cost [185–188]. Therefore, the development of cost-effective multi-sensor fusion algorithms with reasonable reliability and accuracy is critical to the development of MFI.

4.3.3 Integrated Robot System

Another technology applicable to future building sector inspection is to develop a robot system to inspect through walls/floors/ceilings and create high quality assurance standards to avoid argument between inspector and contractor.

According to above review, an A-CONQUAS system, named Quicabot, as shown in Figure 4.10 is proposed, as illustrated in Figure 1.1. It is named QuicaBot [189, 190], an abbreviation for quality inspection and assessment robot. It automatically moves to scan a room in minutes, using color camera, thermal imaging camera and laser scanner to assess architectural defects, including but not limited to cracks and

FIGURE 4.10: Quicabot v1 and selected sensors [10]

uneven surfaces. The system integrates sensing functions into a mobile robot. A thermal camera with pan-tilt plus heating source is used for crack and hollowness detection. An laser scanner with an electronic inclinometer is used for measurement of surface alignment and evenness. The QuicaBot system also allows collaboration of multiple robots to split tasks to increase efficiency. It can perform extensive automated building inspection work through cloud server. The assessor is able to request assessment report from the cloud server and perform a dynamic assessment on a tablet or smart-phone, it also allows multiple assessors to view the results simultaneously. Figure 4.11 shows some typical instances of the assessment results for image-based defects, which are consistent with the ground truth. Note that the main contribution of the author to QuicaBot exists in developing online deep transfer learning algorithms to process the pictures captured by the robot system, to detect the defects. The navigation and control of the robot system are mainly done by others.

Aim to automate inspection, QuicaBot is designed and developed with simultaneous localization and mapping (SLAM) for navigation and positioning, localization in centimetre-level precision, with multiple control speeds to adjust. It equipped with a laser scanner and an inclinometer to check alignment and evenness, a colour camera for crack detecting, and a thermal camera for hollow assessment. The inspection speed is approximately 500 images per minute. Its limitation is that

FIGURE 4.11: Assessment results of QuicaBot for image-based defects

this method requires a lot of data; more meaningful data help improve recognition accuracy. For new applications, whereby only limited data is available, the accuracy might not be guaranteed. A one-day periodic inspection conducted with two assessors can be performed by QuicaBot in half a day; moreover, the QuicaBot can run for 36 hours after two hours of charging, while human can only work for 8-10 hours in a day. Hence, QuicaBot enables more accurate and consistent high quality inspections while reducing the labour and time.

The sharing and storage advancement of big data and information theory has benefited predictive work. Construction quality assessment data can be saved, retrieved and accessed globally through integration services. This enables comparison of inspection results via different approaches, for the same building or similar buildings around the world.

4.4 Coating Condition Assessment - Work in Singapore Polytechnic

Protective coatings are the main route used to prevent corrosion of marine and offshore structures. Inspection of the protective coating is a key issue for asset management. Traditional visual inspection methods are time-consuming and labor-intensive. Hence, in this chapter, efforts were made to develop an automatic or

FIGURE 4.12: Instance-aware semantic segmentation [11]

semi-automatic coating inspection and evaluation system for corrosion management. The system developed is able to quickly and accurately screen and evaluate coating conditions, it can be used as a screening tool for the identification and classification of coating failures.

An image-based coating inspection system is developed for coating breakdown and corrosion (CBC) assessment [16, 18]. The system makes it easy for surveyers to make objective judgments about the CBC's assessment. It improves the effectiveness and reliability of coating inspections and reduces the time and labor required. The system consists of five different phases: ROI recommendations, feature extraction, feature learning, classification, feature detection and CBC measurement.

Three types of CBCs need to be detected, including surface CBC, edge CBC and non-coating failure. In this study, TLCAF network [9] was used to learn CBC features. Bbox is used for ROI proposals, VGG19 model for CNN feature extraction, and SVM for feature classification. The predicted CBC region is then reconstructed for instance aware segmentation. Here, active segmentation [191] is used for background removal. For CBC measurements, hue, saturation, and value (HSV) color space is used for color feature learning. The detected CBC pixel points are used as seeds to facilitate CBC active segmentation.

		Accuracy: 89.54%		
Output Class	edges/welds CBC	87.0% 703	2.6% 24	10.4% 73
	none CBC	2.6% 21	94.8% 876	4.1% 29
	surface CBC	10.4% 84	2.6% 24	85.5% 603
		edges/welds CBC	none CBC	surface CBC
		Target Class		

FIGURE 4.13: Feature prediction results [11]

Instance-aware segmentation is challenging since it requires proper detection of all objects in the image, as well as accurate segmentation of each instance. In this study, CBC detection and segmentation were integrated for CBC instance segmentation. Figure 4.12 shows the typical process of CBC evaluation. Figure 4.12(b) visualizes learned three types of CBC features. The predicted CBC ROIs were reconstructed for active segmentation [191] and CBC grading. Figure 4.12 (d) and (e) demonstrate an example of edge CBC detection and active segmentation. The CBC measurement results are consistent with the ground truth stated in [192].

In this study, 1900 images were labeled for feature extraction; 12,184 features were extracted and divided into three types, namely edge CBC, non-CBC and surface CBC. Randomly select 20% of the total function for verification. Figure 4.13 shows a total verification accuracy of 89.54%. Batch processing module is integrated in the system for autonomous report generation. Every 500 images can be processed in one hour. The study provides a comprehensive automated CBC detection system for marine and offshore industries. A coating condition assessment report is automatically generated according to IACS recommendation 87 and IMO recommendations [193].

4.5 Summary

In this chapter, building quality assessment methods and robotic systems are presented. In order to solve the limitations of existing methods, the online transfer learning for convolution activation features (online-TLCAF) method is proposed

and incorporated into the A-CONQUARS robotic system, called the quality inspection and evaluation robot (Quicabot), for detecting defects. This method has good accuracy in identifying various types of defects including, but not limited to, cracks, finishing defect, corrosion and hollowness.

The investigation in this chapter contributes to automation of quality assessment for large construction sites, which can significantly speed up the quality inspection after construction. The method developed delivers consistent results and high-accuracy assessment at a high level of standard, the automated inspection system automatically processes images at 500 images per minute, which reduces time and increases efficiency for economic benefit.

Chapter 5

Conclusions and Future Work

In this chapter, the major contributions of this thesis are summarized, the potential applications of the technology developed and several directions for future work are discussed.

5.1 Conclusions

The work in this thesis is an important part of a large project to design and develop user-friendly, versatile, efficient and cost-effective robotic systems for automated assessment of construction quality. The proposed A-CONQUARS system is capable of inspecting cracks, finishing defects, hollowness, uniformity and alignment. It includes mobile robots for mapping and navigation, a thermal imaging camera with heater for hollow detection, a color camera for image-based defect detection, and lidar and inclinometer for alignment and evenness inspection. The key focus of this thesis is the development of advanced algorithms for detecting image-based defects, including cracks, finishing defects and hollowness, which will be incorporated into A-CONQUARS to automatically check construction quality.

In this thesis, an online deep transfer learning methodology, i.e. online transfer learning for convolution activation features (online-TLCAF), was developed for detection and recognition of image-based defects in a building. The model and algorithms established displayed superior performance for detection of various defects,

including crack, corrosion, hollowness, etc; they were incorporated into image-based real-time post-construction quality assessment system, which demonstrated significant improvements in accuracy and efficiency for the quality assessment. The novelties of this study are as follows:

- TLCAF and A-TLCAF networks were proposed for image-based defect detection. Compared with the benchmark methods [128, 129], the proposed feature extraction method achieves highest detection rate and the best discriminating power (DP).
- An online-TLCAF network proposed in this work provides incremental learning for high-dimensional dynamic image/video streams. Compared with TLCAF, online-TLCAF has two improvements: 1). The ROI proposal network is replaced by an automated object proposal to eliminate the need for ROI labelling work; 2). the linear classifier in TLCAF is replaced by an online learning system.
- Hollowness beneath the tiles of different materials were studied by active thermography analysis. A-TLCAF enables effective analysis of the captured thermal images to detect the hollowness, in terms of its location, size and shape.
- A robotic platform was designed and developed, which demonstrates the ability to integrate control, sensing and drive, to achieve an intelligent autonomous quality assessment system.
- Compared with traditional manual inspection, the system is suitable for large-area inspection, improving safety, reducing costs, improving the efficiency and reliability of auto-defect assessment, making the results more consistent and resulting in economic benefits. This is the first attempt to apply deep learning techniques to building quality assessment of the construction industry.

Compared to shallow structures, online DTL provides greater flexibility in extracting advanced features and has been proven to promote a variety of scientific and engineering issues. Extensive experiments are validated in CONQUAS room, constructed test bed and our self-built data-set. We illustrate the power of our framework by deriving various learning algorithms. The new automatic evaluation

method is satisfactory for the evaluation various image-based defects. The defect inspection problem is an example of a complex prediction task. We show the successful application of the framework in construction industry. It is of interests not only to researchers working on general-purpose automatized application, but also to those on automated building quality assessment.

5.2 Future Work

Automation of building quality assessment at large construction sites, with consistent results and greater accuracy, helps inspectors reduce time and labor cost for economic benefits. In this thesis, an online-TLCAF methodology was developed, with focus on the inspection of construction projects; an introduction of the current status of automated building quality assessment was also provided. This provides valuable insights for A-CONQUARS and encourages future research. This section discusses potential solutions based on upcoming robotics and related field improvements to gain a deeper understanding of future trends in automated construction quality assessment.

Surface measurement, defect detection and mobile robots have recently been reviewed. The fusion of IRT, mobile robot and laser scanner as building diagnostic tools were studied. The conclusion is that although IRT and laser scanners are very meaningful for Quicabot in the construction industry, however, there is a long way to go to achieve efficient, high-precision, fully automated construction inspection solutions.

Multi-sensor fusion and integration is a big challenge for the remaining technical problems applied in the actual building quality assessment system. The three-dimensional reconstruction of the environment and the image index to indicate the location of defects will help visualize defects and integrate the entire system in the cloud to promote collaboration between different parties. At the same time, data collection is an ongoing process; more meaningful data help improve defect recognition accuracy. Multi-model AI can help improve the accuracy of defect detection system and make the system more powerful and powerful.

Fuzzy online TLCAF will further increase the speed of operation. Knowledge graph will incorporate heterogeneous signals for collaborative knowledge discovery

to improve decision making. In addition, light robots require cloud computing to achieve high intelligence. Cloud-based frameworks help with computing, sharing storage and knowledge, and dynamically allocating resources. Recent research has shown that cloud servers can make robots smarter by offloading computing loads and using the Internet to provide services on demand. This involves integrating cloud computing technology into the robot and using the collected data to assess the quality of the building in the cloud. Intelligent mobile cloud robots have the potential to automate building quality assessments, and the use of such robots can greatly reduce the human resources required for missions. The integrated intelligent robotic system will be able to capture defects through existing sensors by using defects in the cloud-based framework to detect defects, mark defects in the map and generate reports. Cloud-based robots will be able to navigate, index defect locations and collaborate with assessors.

List of Author's Patents and Publications¹

Technology Disclosure (TD) & Patent

1. Y Zhen, Z Cai, L Liu, E Tan, CNN-based Automatic Coating Inspection method, TD, Singapore Polytechnic, 2019.
2. Y Zhen, Z Cai, L Liu, E Tan, X J Yin, Corrosion assessment method based on image and system thereof, TD, Singapore Polytechnic, 2017.
3. E Kayacan, I M Chen, L K Tiong, L Liu, V Maruvanchery, R J Yan, Building defect assessment method based on image and system thereof, TD, NTU, 2016.
4. A Causo, L Liu, I M Chen, S H Yeo, CPR compression performance measurement device and application, TD, NTU, Oct. 2015.
5. I M Chen, L Liu, B Li, M Sung, T Goh, W Soh, D Fung, Y P Ooi, S Weng, Robot-facilitated interaction for children with autism spectrum disorder, copyright, NTU, April. 2015.
6. I M Chen, S H Yeo, A Causo, L Liu, Q Yuan. Embedded wireless stroke rehabilitation system employing handheld device which includes short range radio transceivers. TD, NTU, 2012
7. W Ma, W Zheng, L Liu. Integrated capillary electrophoresis chip scanning and analyzing system, patent, publication no. CN1821747, application no. 200610025229.6.

¹The superscript * indicates joint first authors

Journal Articles

1. Lili Liu, I-Ming Chen. OnlineTLCAF: Online learning of deep transferred features, ICRA 2020 and RA-letter, submitted, 2020.
2. Lili Liu, I-Ming Chen. Building machines that learn like human - Online deep transfer learning review, IEEE transactions on automation science and engineering (T-ASE), submitted, 2019.
3. Lili Liu, Estee Tan, Zhi Qiang Cai, Xi Jiang Yin, and Yongda Zhen. CNN-based Automatic Coating Inspection System, Advances in Science, Technology and Engineering Systems Journal, 2018.
4. Lili Liu, Rui-Jun Yan, Varun Maruvanchery, Erdal Kayacan, I-Ming Chen, and Lee Kong Tiong. Transfer learning on convolutional activation feature as applied to a building quality assessment robot. International Journal of Advanced Robotic Systems, 14(3):1729881417712620, 2017.
5. Lili Liu, Wenli Ma, Wenjuan Yao, and Wenling Zheng. Research progress in integrated capillary electrophoresis chips, Beijing Biomedical Engineering, 25(3):316-320, 2006.
6. Lili Liu, Wenjuan Yao, Wenli Ma, and Wenling Zheng. Integrated capillary electrophoresis chip scanning and analyzing system based on dsp and ccd. Chinese Journal of Sensors and Actuators, 19(2):341-345, 2006.

Conference Proceedings

1. Lili Liu, Estee Tan, Zhi Qiang Cai, and Yongda Zhen. Deep learning for coating condition assessment with active perception, The 2nd International Conference on Big Data and Artificial Intelligence (BDAI), Guangzhou, China, 2019.
2. Lili Liu, Estee Tan, Yongda Zhen, Xi Jiang Yin, and Zhi Qiang Cai. AI facilitated coating corrosion assessment system for productivity enhancement. In 2018 13th IEEE Conference on Industrial Electronics and Applications (ICIEA), pages 606-610. IEEE, 2018.

3. Lili Liu, Estee Tan, Zhi Qiang Cai, Yongda Zhen, and Xi Jiang Yin. An integrated coating inspection system for marine and offshore corrosion management. In 2018 15th International Conference on Control, Automation, Robotics and Vision (ICARCV), pages 1531-1536. IEEE, 2018.
4. Lili Liu, Estee Tan, Kevin Lim, Xi Jiang Yin, Zhi Qiang Cai, Yongda Zhen, and Hai Gu. Artificial intelligence for vision-based inspection for marine and offshore assets, IADC drilling middle east 2018 conference & exhibition, Dubai, 2018.
5. Estee Tan, Lili Liu, Yongda Zhen, Xi Jiang Yin, and Zhi Qiang Cai. Non-destructive detection of corrosion under coatings by active infrared thermography, Far East NDT New Technology & Application Forum, Xianmen, China, 2018.
6. Lili Liu, Estee Tan, Kelvin Chee Quan Lim, Joo En Ng, Zhi Qiang Cai, Nengfu Tao, Yongda Zhen, and Xi Jiang Yin. Fast inspection of structure defect for risk assessment by active thermography, 42nd Conference on Our World in Concrete & Structures, Singapore, 2017.
7. Rui-Jun Yan, Chin Leong Low, Jinjun Duan, Lili Liu, Erdal Kayacan, I-Ming Chen, and Robert Tiong. Development of a novel post-construction quality assessment robot system. In Control, Automation, Robotics and Vision (ICARCV), 2016 14th International Conference on, pages 1-6. IEEE, 2016.
8. Lili Liu, I-Ming Chen, Erdal Kayacan, Lee Kong Tiong, and Varun Maruvanchery. Automated construction quality assessment: A review. In Mechatronics and its Applications (ISMA), 2015 10th International Symposium on, pages 1-6. IEEE, 2015.
9. Lili Liu, Bingbing Li, I-Ming Chen, Tze Jui Goh, and Min Sung. Interactive robots as social partner for communication care. In 2014 IEEE International Conference on Robotics and Automation (ICRA), pages 2231-2236. IEEE, 2014.
10. Albert Causo, Lili Liu, Ganesh Krishnasamy, Song Huat Yeo, and I-Ming Chen. Integration of wireless wearable sensors and mobile computing with

- cloud-based service for patient rehabilitation monitoring. In Proceedings of International Conference on Intelligent Unmanned Systems, volume 8, 2012.
11. Zhiqiang Luo, Weiting Yang, Zhong Qiang Ding, Lili Liu, I-Ming Chen, Song Huat Yeo, Keck Voon Ling, and Henry Been-Lirn Duh. “left arm up!” interactive yoga training in virtual environment. In 2011 IEEE Virtual Reality Conference, pages 261-262. IEEE, 2011.

Bibliography

- [13] Iti Chaturvedi, Yew-Soon Ong, and Rajesh Vellore Arumugam. Deep transfer learning for classification of time-delayed gaussian networks. *Signal Processing*, 110:250–262, 2015.
- [14] Long Wen, Liang Gao, and Xinyu Li. A new deep transfer learning based on sparse auto-encoder for fault diagnosis. *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, 2017.
- [15] CL Philip Chen and Zhulin Liu. Broad learning system: An effective and efficient incremental learning system without the need for deep architecture. *IEEE transactions on neural networks and learning systems*, 29(1):10–24, 2017.
- [11] Lili LIU, Estee Tan, Zhi Qiang Cai, Xi Jiang Yin, and Yongda Zhen. Cnn-based automatic coating inspection system. *Advances in Science, Technology and Engineering Systems Journal*, 3(6):469–478, 2018.
- [9] Lili Liu, Rui-Jun Yan, Varun Maruvanchery, Erdal Kayacan, I-Ming Chen, and Lee Kong Tiong. Transfer learning on convolutional activation feature as applied to a building quality assessment robot. *International Journal of Advanced Robotic Systems*, 14(3):1729881417712620, 2017.
- [16] Lili Liu, Estee Tan, Yongda Zhen, Xi Jiang Yin, and Zhi Qiang Cai. Ai-facilitated coating corrosion assessment system for productivity enhancement. In *2018 13th IEEE Conference on Industrial Electronics and Applications (ICIEA)*, pages 606–610. IEEE, 2018.
- [17] Lili Liu, Estee Tan, Zhi Qiang Cai, Yongda Zhen, and Xi Jiang Yin. An integrated coating inspection system for marine and offshore corrosion management. In *The 15th International Conference on Control, Automation, Robotics and Vision (ICARCV 2018)*. IEEE, 2018.

- [18] Lili Liu, Estee Tan, Kevin Lim, Yiheng Ma, Xi Jiang Yin, Zhi Qiang Cai, and Yongda Zhen. Artificial intelligence for vision-based inspection for marine and offshore assets. In *IADC drilling middle east 2018 conference and exhibition*. International Association of Drilling Contractors, USA, 2018.
- [19] Lili Liu, I-Ming Chen, Erdal Kayacan, Lee Kong Tiong, and Varun Maruvanchery. Automated construction quality assessment: A review. In *Mechanics and its Applications (ISMA), 2015 10th International Symposium on*, pages 1–6. IEEE, 2015.
- [20] Lili Liu, Estee Tan, Kelvin Chee Quan Lim, Joo En Ng, Zhi Qiang Cai, Nengfu Tao, Yongda Zhen, and XiJiang Yin. Fast inspection of structure defect for risk assessment by active thermography. In *42nd Conference on Our World in Concrete & Structures*, pages 1–6, Singapore, 2017.
- [21] Estee Tan, Lili LIU, Yongda Zhen, Xijiang Yin, and Zhi Qiang Cai. Non-destructive detection of corrosion under coatings by active infrared thermography. In *Far East NDT New Technology & Application Forum*, pages 1–6, Xianmen, China, 2018.
- [22] Lili LIU, Rui-Jun Yan, Varun Maruvanchery, Erdal Kayacan, I-Ming Chen, and Lee Kong Tiong. Automated inspection of building quality: status and prospects. *International journal of robotics and automation (IJRA)*, submitted, 2017.
- [23] Rui-Jun Yan, Chin Leong Low, Jinjun Duan, Lili Liu, Erdal Kayacan, I-Ming Chen, and Robert Tiong. Development of a novel post-construction quality assessment robot system. In *Control, Automation, Robotics and Vision (ICARCV), 2016 14th International Conference on*, pages 1–6. IEEE, 2016.
- [24] Jordan Guerguiev, Timothy P Lillicrap, and Blake A Richards. Towards deep learning with segregated dendrites. *ELife*, 6:e22901, 2017.
- [25] Luke Mastin. Neurons & synapses, 2018. URL http://www.human-memory.net/brain_neurons.html.
- [1] M Ahmadi, H Naderpour, and A Kheyroddin. Utilization of artificial neural networks to prediction of the capacity of ccft short columns subject to short

- term axial load. *Archives of civil and mechanical engineering*, 14(3):510–517, 2014. [8](#)
- [26] Wenling Zheng and Wenli Ma. Thinking mechanism. *Technology Review*, 13(9505):16–19, 1995.
- [27] David H Hubel and Torsten N Wiesel. Receptive fields of single neurones in the cat’s striate cortex. *The Journal of physiology*, 148(3):574–591, 1959.
- [28] Jeff Hawkins and Sandra Blakeslee. *On intelligence: How a new understanding of the brain will lead to the creation of truly intelligent machines*. Macmillan, 2007.
- [29] Zoya Bylinskii, Tilke Judd, Ali Borji, Laurent Itti, Frédo Durand, Aude Oliva, and Antonio Torralba. Mit saliency benchmark, 2015.
- [30] Ming-Ming Cheng, Niloy J Mitra, Xiaolei Huang, Philip HS Torr, and Shi-Min Hu. Global contrast based salient region detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 37(3):569–582, 2015.
- [31] Ming-Ming Cheng, Ziming Zhang, Wen-Yan Lin, and Philip Torr. Bing: Binarized normed gradients for objectness estimation at 300fps. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 3286–3293, 2014.
- [32] Pengpeng Liang, Yu Pang, Chunyuan Liao, Xue Mei, and Haibin Ling. Adaptive objectness for object tracking. *IEEE Signal Process. Lett.*, 23(7):949–953, 2016.
- [33] R Sai Srivatsa and R Venkatesh Babu. Salient object detection via objectness measure. In *Image Processing (ICIP), 2015 IEEE International Conference on*, pages 4481–4485. IEEE, 2015.
- [34] Huiying Liu, Shuqiang Jiang, Qingming Huang, Changsheng Xu, and Wen Gao. Region-based visual attention analysis with its application in image browsing on small displays. In *Proceedings of the 15th ACM international conference on Multimedia*, pages 305–308. ACM, 2007.
- [2] Ali Borji, Ming-Ming Cheng, Huaizu Jiang, and Jia Li. Salient object detection: A survey. arxiv preprint. *arXiv preprint arXiv:1411.5878*, 2(4), 2014. [12](#), [13](#)

- [35] Ali Borji, Dicky N Sihite, and Laurent Itti. Quantitative analysis of human-model agreement in visual saliency modeling: A comparative study. *IEEE Transactions on Image Processing*, 22(1):55–69, 2013.
- [36] Ali Borji, Ming-Ming Cheng, Huaizu Jiang, and Jia Li. Salient object detection: A benchmark. *IEEE transactions on image processing*, 24(12):5706–5722, 2015.
- [37] Gang Hua, Zicheng Liu, Zhengyou Zhang, and Ying Wu. Iterative local-global energy minimization for automatic extraction of objects of interest. *IEEE transactions on pattern analysis and machine intelligence*, 28(10):1701–1706, 2006.
- [38] Mohand Said Allili and Djemel Ziou. Object of interest segmentation and tracking by using feature selection and active contours. In *2007 IEEE Conference on Computer Vision and Pattern Recognition*, pages 1–8. IEEE, 2007.
- [39] Laurent Itti, Christof Koch, and Ernst Niebur. A model of saliency-based visual attention for rapid scene analysis. *IEEE Transactions on pattern analysis and machine intelligence*, 20(11):1254–1259, 1998.
- [40] Dorin Comaniciu and Peter Meer. Mean shift: A robust approach toward feature space analysis. *IEEE Transactions on pattern analysis and machine intelligence*, 24(5):603–619, 2002.
- [41] Pedro F Felzenszwalb and Daniel P Huttenlocher. Efficient graph-based image segmentation. *International journal of computer vision*, 59(2):167–181, 2004.
- [42] Radhakrishna Achanta, Appu Shaji, Kevin Smith, Aurelien Lucchi, Pascal Fua, Sabine Süsstrunk, et al. Slic superpixels compared to state-of-the-art superpixel methods. *IEEE transactions on pattern analysis and machine intelligence*, 34(11):2274–2282, 2012.
- [43] Alex Levinstein, Adrian Stere, Kiriakos N Kutulakos, David J Fleet, Sven J Dickinson, and Kaleem Siddiqi. Turbopixels: Fast superpixels using geometric flows. *IEEE transactions on pattern analysis and machine intelligence*, 31(12):2290–2297, 2009.

- [44] Jonathan Long, Evan Shelhamer, and Trevor Darrell. Fully convolutional networks for semantic segmentation. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 3431–3440, 2015.
- [45] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 770–778, 2016.
- [46] Karen Simonyan and Andrew Zisserman. Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556*, 2014.
- [47] Bogdan Alexe, Thomas Deselaers, and Vittorio Ferrari. What is an object? In *Computer Vision and Pattern Recognition (CVPR), 2010 IEEE Conference on*, pages 73–80. IEEE, 2010.
- [48] Bogdan Alexe, Thomas Deselaers, and Vittorio Ferrari. Measuring the objectness of image windows. *IEEE transactions on pattern analysis and machine intelligence*, 34(11):2189–2202, 2012.
- [49] Ian Endres and Derek Hoiem. Category independent object proposals. In *European Conference on Computer Vision*, pages 575–588. Springer, 2010.
- [50] Ian Endres and Derek Hoiem. Category-independent object proposals with diverse ranking. *IEEE transactions on pattern analysis and machine intelligence*, 36(2):222–234, 2014.
- [51] Jasper RR Uijlings, Koen EA Van De Sande, Theo Gevers, and Arnold WM Smeulders. Selective search for object recognition. *International journal of computer vision*, 104(2):154–171, 2013.
- [52] Ziming Zhang, Jonathan Warrell, and Philip HS Torr. Proposal generation for object detection using cascaded ranking svms. In *Computer Vision and Pattern Recognition (CVPR), 2011 IEEE Conference on*, pages 1497–1504. IEEE, 2011.
- [53] Ross Girshick. Fast r-cnn. In *Proceedings of the IEEE international conference on computer vision*, pages 1440–1448, 2015.

- [54] Ross Girshick, Jeff Donahue, Trevor Darrell, and Jitendra Malik. Rich feature hierarchies for accurate object detection and semantic segmentation. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 580–587, 2014.
- [55] Shaoqing Ren, Kaiming He, Ross Girshick, and Jian Sun. Faster r-cnn: Towards real-time object detection with region proposal networks. pages 91–99, 2015.
- [3] Joseph Redmon, Santosh Divvala, Ross Girshick, and Ali Farhadi. You only look once: Unified, real-time object detection. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 779–788, 2016. [14](#), [15](#), [19](#), [20](#), [47](#), [48](#)
- [4] Joseph Redmon and Ali Farhadi. Yolo9000: better, faster, stronger. *arXiv preprint*, 2017. [14](#), [15](#), [19](#), [20](#), [47](#), [48](#)
- [56] Mark Everingham, SM Ali Eslami, Luc Van Gool, Christopher KI Williams, John Winn, and Andrew Zisserman. The pascal visual object classes challenge: A retrospective. *International journal of computer vision*, 111(1): 98–136, 2015.
- [57] Mark Everingham, Luc Van Gool, Christopher KI Williams, John Winn, and Andrew Zisserman. The pascal visual object classes (voc) challenge. *International journal of computer vision*, 88(2):303–338, 2010.
- [58] Mark Everingham, Luc Van Gool, Christopher KI Williams, John Winn, and Andrew Zisserman. The pascal visual object classes challenge 2007 (voc2007) results. 2007.
- [59] Subhransu Maji, Alexander C Berg, and Jitendra Malik. Classification using intersection kernel support vector machines is efficient. In *Computer Vision and Pattern Recognition, 2008. CVPR 2008. IEEE Conference on*, pages 1–8. IEEE, 2008.
- [60] Andrea Vedaldi and Andrew Zisserman. Efficient additive kernels via explicit feature maps. *IEEE transactions on pattern analysis and machine intelligence*, 34(3):480–492, 2012.

- [61] Dengsheng Lu and Qihao Weng. A survey of image classification methods and techniques for improving classification performance. *International journal of Remote sensing*, 28(5):823–870, 2007.
- [62] Siddhartha Sankar Nath, Girish Mishra, Jajnyaseni Kar, Sayan Chakraborty, and Nilanjan Dey. A survey of image classification methods and techniques. In *Control, Instrumentation, Communication and Computational Technologies (ICCICCT), 2014 International Conference on*, pages 554–557. IEEE, 2014.
- [63] Pooja Kamavisdar, Sonam Saluja, and Sonu Agrawal. A survey on image classification approaches and techniques. *International Journal of Advanced Research in Computer and Communication Engineering*, 2(1):1005–1009, 2013.
- [64] Jia Deng, Wei Dong, Richard Socher, Li-Jia Li, Kai Li, and Li Fei-Fei. Imagenet: A large-scale hierarchical image database. In *Computer Vision and Pattern Recognition, 2009. CVPR 2009. IEEE Conference on*, pages 248–255. Ieee, 2009.
- [65] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E Hinton. Imagenet classification with deep convolutional neural networks. In *Advances in neural information processing systems*, pages 1097–1105, 2012.
- [66] Financial Times. Machines ‘beat humans’ for a growing number of tasks, 2018. URL <https://www.ft.com/content/4cc048f6-d5f4-11e7-a303-9060cb1e5f44>.
- [67] Olga Russakovsky, Jia Deng, Hao Su, Jonathan Krause, Sanjeev Satheesh, Sean Ma, Zhiheng Huang, Andrej Karpathy, Aditya Khosla, Michael Bernstein, et al. Imagenet large scale visual recognition challenge. *International Journal of Computer Vision*, 115(3):211–252, 2015.
- [68] Jeremy Howard and Sebastian Ruder. Universal language model fine-tuning for text classification. *arXiv preprint arXiv:1801.06146*, 2018.
- [69] Saining Xie, Tianbao Yang, Xiaoyu Wang, and Yuanqing Lin. Hyper-class augmented and regularized deep learning for fine-grained image classification. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 2645–2654, 2015.

- [70] C Lawrence Zitnick and Piotr Dollár. Edge boxes: Locating object proposals from edges. In *European conference on computer vision*, pages 391–405. 2014.
- [71] Pablo Arbeláez, Jordi Pont-Tuset, Jonathan T Barron, Ferran Marques, and Jitendra Malik. Multiscale combinatorial grouping. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 328–335, 2014.
- [72] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Spatial pyramid pooling in deep convolutional networks for visual recognition. pages 346–361, 2014.
- [73] Jan Hosang, Rodrigo Benenson, Piotr Dollár, and Bernt Schiele. What makes for effective detection proposals? *IEEE transactions on pattern analysis and machine intelligence*, 38(4):814–830, 2016.
- [5] Sebastian Ruder. Transfer learning: Machine learnings next frontier, 2017. [23](#)
- [6] Sinno Jialin Pan, Qiang Yang, et al. A survey on transfer learning. *IEEE Transactions on knowledge and data engineering*, 22(10):1345–1359, 2010. [23](#), [26](#), [28](#), [39](#)
- [74] Emilio Soria Olivas. *Handbook of Research on Machine Learning Applications and Trends: Algorithms, Methods, and Techniques: Algorithms, Methods, and Techniques*. IGI Global, 2009.
- [75] Lisa Torrey and Jude Shavlik. Transfer learning. pages 242–264, 2010.
- [76] Gert Cauwenberghs and Tomaso Poggio. Incremental and decremental support vector machine learning. In *Advances in neural information processing systems*, pages 409–415, 2001.
- [77] Battista Biggio, Iginio Corona, Blaine Nelson, Benjamin IP Rubinstein, Davide Maiorca, Giorgio Fumera, Giorgio Giacinto, and Fabio Roli. Security evaluation of support vector machines in adversarial environments. In *Support Vector Machines Applications*, pages 105–153. Springer, 2014.
- [78] Yanyun Lu, Khaled Boukharouba, Jacques Boonært, Anthony Fleury, and Stéphane Lecoecuche. Application of an incremental svm algorithm for on-line

- human recognition from video surveillance using texture and color features. *Neurocomputing*, 126:132–140, 2014.
- [79] Antoine Bordes, Seyda Ertekin, Jason Weston, and Léon Bottou. Fast kernel classifiers with online and active learning. *Journal of Machine Learning Research*, 6(Sep):1579–1619, 2005.
- [80] Cho-Jui Hsieh, Si Si, and Inderjit Dhillon. A divide-and-conquer solver for kernel support vector machines. In *International Conference on Machine Learning*, pages 566–574, 2014.
- [81] Zhaowei Cai, Longyin Wen, Zhen Lei, Nuno Vasconcelos, and Stan Z Li. Robust deformable and occluded object tracking with dynamic graph. *IEEE Transactions on Image Processing*, 23(12):5497–5509, 2014.
- [82] Amir Saffari, Christian Leistner, Jakob Santner, Martin Godec, and Horst Bischof. On-line random forests. In *Computer Vision Workshops (ICCV Workshops), 2009 IEEE 12th International Conference on*, pages 1393–1400. IEEE, 2009.
- [83] Federico Pernici and Alberto Del Bimbo. Object tracking by oversampling local features. *IEEE transactions on pattern analysis and machine intelligence*, 36(12):2538–2551, 2014.
- [84] Harry Zhang. The optimality of naive bayes. *AA*, 1(2):3, 2004.
- [85] Vangelis Metsis, Ion Androutsopoulos, and Georgios Paliouras. Spam filtering with naive bayes-which naive bayes? In *CEAS*, volume 17, pages 28–69. Mountain View, CA, 2006.
- [86] SL Ting, WH Ip, and Albert HC Tsang. Is naive bayes a good classifier for document classification? *International Journal of Software Engineering and Its Applications*, 5(3):37–46, 2011.
- [87] Nan-Ying Liang, Guang-Bin Huang, Paramasivan Saratchandran, and Narasimhan Sundararajan. A fast and accurate online sequential learning algorithm for feedforward networks. *IEEE Transactions on neural networks*, 17(6):1411–1423, 2006.

- [88] Jiexiong Tang, Chenwei Deng, and Guang-Bin Huang. Extreme learning machine for multilayer perceptron. *IEEE transactions on neural networks and learning systems*, 27(4):809–821, 2016.
- [89] Jiexiong Tang, Chenwei Deng, Guang-Bin Huang, and Baojun Zhao. Compressed-domain ship detection on spaceborne optical image using deep neural network and extreme learning machine. *IEEE Transactions on Geoscience and Remote Sensing*, 53(3):1174–1185, 2015.
- [90] Zeynep Akata, Florent Perronnin, Zaid Harchaoui, and Cordelia Schmid. Good practice in large-scale learning for image classification. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 36(3):507–520, 2014.
- [91] Michael Sapienza, Fabio Cuzzolin, and Philip HS Torr. Learning discriminative space-time action parts from weakly labelled videos. *International journal of computer vision*, 110(1):30–47, 2014.
- [92] Gail A Carpenter, Stephen Grossberg, and David B Rosen. Fuzzy art: Fast stable learning and categorization of analog patterns by an adaptive resonance system. *Neural networks*, 4(6):759–771, 1991.
- [93] Gail A Carpenter and Ah-Hwee Tan. Rule extraction, fuzzy artmap, and medical databases. Technical report, Boston University Center for Adaptive Systems and Department of Cognitive , 1993.
- [94] Ah-Hwee Tan. Cascade artmap: Integrating neural computation and symbolic knowledge processing. *IEEE Transactions on Neural Networks*, 8(2):237–250, 1997.
- [95] Zhepei Wei, Di Wang, Ming Zhang, Ah-Hwee Tan, Chunyan Miao, and You Zhou. Autonomous agents in snake game via deep reinforcement learning. In *2018 IEEE International Conference on Agents (ICA)*, pages 20–25. IEEE, 2018.
- [96] Volodymyr Mnih, Koray Kavukcuoglu, David Silver, Andrei A Rusu, Joel Veness, Marc G Bellemare, Alex Graves, Martin Riedmiller, Andreas K Fidjeland, Georg Ostrovski, et al. Human-level control through deep reinforcement learning. *Nature*, 518(7540):529, 2015.

- [97] David Silver, Aja Huang, Chris J Maddison, Arthur Guez, Laurent Sifre, George Van Den Driessche, Julian Schrittwieser, Ioannis Antonoglou, Veda Panneershelvam, Marc Lanctot, et al. Mastering the game of go with deep neural networks and tree search. *nature*, 529(7587):484, 2016.
- [98] Yoh-Han Pao, Gwang-Hoon Park, and Dejan J Sobajic. Learning and generalization characteristics of the random vector functional-link net. *Neurocomputing*, 6(2):163–180, 1994.
- [99] CL Philip Chen and Zhulin Liu. Broad learning system: an effective and efficient incremental learning system without the need for deep architecture. *IEEE transactions on neural networks and learning systems*, 29(1):10–24, 2018.
- [100] Viktor Losing, Barbara Hammer, and Heiko Wersing. Incremental on-line learning: A review and comparison of state of the art algorithms. *Neurocomputing*, 275:1261–1274, 2018.
- [7] Chuanqi Tan, Fuchun Sun, Tao Kong, Wenchang Zhang, Chao Yang, and Chunfang Liu. A survey on deep transfer learning. In *International Conference on Artificial Neural Networks*, pages 270–279. Springer, 2018. [38](#), [39](#)
- [101] Karl Weiss, Taghi M Khoshgoftaar, and DingDing Wang. A survey of transfer learning. *Journal of Big Data*, 3(1):9, 2016.
- [102] Na Li, Huizhen Hao, Qing Gu, Danru Wang, and Xiumian Hu. A transfer learning method for automatic identification of sandstone microscopic images. *Computers & Geosciences*, 103:111–121, 2017.
- [103] Mingsheng Long, Yue Cao, Jianmin Wang, and Michael I Jordan. Learning transferable features with deep adaptation networks. *arXiv preprint arXiv:1502.02791*, 2015.
- [104] Mingsheng Long, Han Zhu, Jianmin Wang, and Michael I Jordan. Deep transfer learning with joint adaptation networks. *arXiv preprint arXiv:1605.06636*, 2016.
- [105] Eric Tzeng, Judy Hoffman, Ning Zhang, Kate Saenko, and Trevor Darrell. Deep domain confusion: Maximizing for domain invariance. *arXiv preprint arXiv:1412.3474*, 2014.

- [106] Hang Chang, Ju Han, Cheng Zhong, Antoine M Snijders, and Jian-Hua Mao. Unsupervised transfer learning via multi-scale convolutional sparse coding for biomedical applications. *IEEE transactions on pattern analysis and machine intelligence*, 40(5):1182–1194, 2018.
- [107] Daniel George, Hongyu Shen, and EA Huerta. Deep transfer learning: A new deep learning glitch classification method for advanced ligo. *arXiv preprint arXiv:1706.07446*, 2017.
- [108] Maxime Oquab, Leon Bottou, Ivan Laptev, and Josef Sivic. Learning and transferring mid-level image representations using convolutional neural networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 1717–1724, 2014.
- [109] Han Zhu, Mingsheng Long, Jianmin Wang, and Yue Cao. Deep hashing network for efficient similarity retrieval. In *AAAI*, pages 2415–2421, 2016.
- [110] Hana Ajakan, Pascal Germain, Hugo Larochelle, François Laviolette, and Mario Marchand. Domain-adversarial neural networks. *arXiv preprint arXiv:1412.4446*, 2014.
- [111] Yaroslav Ganin and Victor Lempitsky. Unsupervised domain adaptation by backpropagation. *arXiv preprint arXiv:1409.7495*, 2014.
- [112] Zelun Luo, Yuliang Zou, Judy Hoffman, and Li F Fei-Fei. Label efficient learning of transferable representations across domains and tasks. In *Advances in Neural Information Processing Systems*, pages 165–177, 2017.
- [113] Eric Tzeng, Judy Hoffman, Trevor Darrell, and Kate Saenko. Simultaneous deep transfer across domains and tasks. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 4068–4076, 2015.
- [114] Eric Tzeng, Judy Hoffman, Kate Saenko, and Trevor Darrell. Adversarial discriminative domain adaptation. In *Computer Vision and Pattern Recognition (CVPR)*, volume 1, page 4, 2017.
- [115] Ah-Hwee Tan. Falcon: A fusion architecture for learning, cognition, and navigation. In *2004 IEEE International Joint Conference on Neural Networks (IEEE Cat. No. 04CH37541)*, volume 4, pages 3297–3302. IEEE, 2004.

- [116] William Gu, Gerald Seet, and Nadia Magnenat-Thalmanna. Perception-link behavior model: Supporting a novel operator interface for a customizable anthropomorphic telepresence robot. *Robotics*, 6(3):16, 2017.
- [117] CL Philip Chen and John Z Wan. A rapid learning and dynamic stepwise updating algorithm for flat neural networks and the application to time-series prediction. *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics)*, 29(1):62–72, 1999.
- [118] Y-H Pao and Yoshiyasu Takefuji. Functional-link net computing: theory, system architecture, and functionalities. *Computer*, 25(5):76–79, 1992.
- [119] Le Zhang and Ponnuthurai N Suganthan. A comprehensive evaluation of random vector functional link networks. *Information sciences*, 367:1094–1105, 2016.
- [120] Chang Xu, Dacheng Tao, and Chao Xu. A survey on multi-view learning. *arXiv preprint arXiv:1304.5634*, 2013.
- [121] Wanli Ouyang, Xiao Chu, and Xiaogang Wang. Multi-source deep learning for human pose estimation. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 2329–2336, 2014.
- [122] Shuo Xiang, Lei Yuan, Wei Fan, Yalin Wang, Paul M Thompson, and Jieping Ye. Multi-source learning with block-wise missing data for alzheimer’s disease prediction. In *Proceedings of the 19th ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 185–193. ACM, 2013.
- [123] Nachiketa Acharya, Nitin Anand Shrivastava, BK Panigrahi, and UC Mohanty. Development of an artificial neural network based multi-model ensemble to estimate the northeast monsoon rainfall over south peninsular india: an application of extreme learning machine. *Climate dynamics*, 43(5-6):1303–1310, 2014.
- [124] Hyeonseob Nam and Bohyung Han. Learning multi-domain convolutional neural networks for visual tracking. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 4293–4302, 2016.
- [125] Yongxin Yang and Timothy M Hospedales. A unified perspective on multi-domain and multi-task learning. *arXiv preprint arXiv:1412.7489*, 2014.

- [12] Cody Coleman, Deepak Narayanan, Daniel Kang, Tian Zhao, Jian Zhang, Luigi Nardi, Peter Bailis, Kunle Olukotun, Chris Ré, and Matei Zaharia. Dawnbench: An end-to-end deep learning benchmark and competition. *Training*, 100(101):102, 2017.
- [126] Mathworks. Pretrained convolutional neural networks, 2018. URL <https://ww2.mathworks.cn/help/deeplearning/ug/pretrained-convolutional-neural-networks.html;jsessionid=de41dc9fba7ecda55cc2250a21be>.
- [8] Alfredo Canziani, Adam Paszke, and Eugenio Culurciello. An analysis of deep neural network models for practical applications. *arXiv preprint arXiv:1605.07678*, 2016. 60
- [127] Christian Szegedy, Sergey Ioffe, Vincent Vanhoucke, and Alexander A Alemi. Inception-v4, inception-resnet and the impact of residual connections on learning. In *Thirty-First AAAI Conference on Artificial Intelligence*, 2017.
- [128] Mohammad R Jahanshahi and Sami F Masri. Adaptive vision-based crack detection using 3d scene reconstruction for condition assessment of structures. *Automation in Construction*, 22:567–576, 2012.
- [129] S Nashat, Azizi Abdullah, and MZ Abdullah. Machine vision for crack inspection of biscuits featuring pyramid detection scheme. *Journal of Food Engineering*, 120:233–247, 2014.
- [130] Ronny Salim Lim, Hung Manh La, Zeyong Shan, and Weihua Sheng. Developing a crack inspection robot for bridge maintenance. In *Robotics and Automation (ICRA), 2011 IEEE International Conference on*, pages 6288–6293. IEEE, 2011.
- [131] Matthew D Zeiler and Rob Fergus. Visualizing and understanding convolutional networks. In *European conference on computer vision*, pages 818–833. Springer, 2014.
- [132] Marina Sokolova, Nathalie Japkowicz, and Stan Szpakowicz. Beyond accuracy, f-score and roc: a family of discriminant measures for performance evaluation. In *Australasian joint conference on artificial intelligence*, pages 1015–1021. 2006.

- [133] Aitor Ibarguren, Jorge Molina, Loreto Susperregi, and Inaki Maurtua. Thermal tracking in mobile robots for leak inspection activities. *Sensors*, 13(10): 13560–13574, 2013.
- [134] Ying-Chieh Chou and Leehter Yao. Automatic diagnostic system of electrical equipment using infrared thermography. In *Soft Computing and Pattern Recognition, 2009. SOCPAR'09. International Conference of*, pages 155–160. IEEE, 2009.
- [135] Jonathan Henrique Efigênio de Oliveira and Walter Fetter Lages. Robotized inspection of power lines with infrared vision. In *Applied Robotics for the Power Industry (CARPI), 2010 1st International Conference on*, pages 1–6. IEEE, 2010.
- [136] Ryan Ahmed, Mohammed El Sayed, S Andrew Gadsden, Jimi Tjong, and Saeid Habibi. Automotive internal-combustion-engine fault detection and classification using artificial neural network techniques. *IEEE Transactions on vehicular technology*, 64(1):21–33, 2015.
- [137] Mohd Anuar Shafi'i and Noraliza Hamzah. Internal fault classification using artificial neural network. In *Power Engineering and Optimization Conference (PEOCO), 2010 4th International*, pages 352–357. IEEE, 2010.
- [138] Baoshu Li, Xiaohui Zhu, Shutao Zhao, and Wendong Niu. Hv power equipment diagnosis based on infrared imaging analyzing. In *Power System Technology, 2006. PowerCon 2006. International Conference on*, pages 1–4. IEEE, 2006.
- [139] Abolfazl Rahmani, Javad Haddadnia, and Omid Seryasat. Intelligent fault detection of electrical equipment in ground substations using thermo vision technique. In *Mechanical and Electronics Engineering (ICMEE), 2010 2nd International Conference on*, volume 2, pages V2–150. IEEE, 2010.
- [140] Carlos A Laurentys Almeida, Antônio P Braga, Sinval Nascimento, Vinicius Paiva, Hêlvio JA Martins, Rodolfo Torres, and Walmir M Caminhas. Intelligent thermographic diagnostic applied to surge arresters: a new approach. *IEEE Transactions on Power Delivery*, 24(2):751–757, 2009.

- [141] Soib Taib, Mohd Shawal Jadin, and Shahid Kabir. Thermal imaging for enhancing inspection reliability: detection and characterization. In *Infrared Thermography*. InTech, 2012.
- [142] Prateek Prasanna, Kristin J Dana, Nenad Gucunski, Basily B Basily, Hung Manh La, Ronny Salim Lim, and Hooman Parvardeh. Automated crack detection on concrete bridges. *IEEE Trans. Automation Science and Engineering*, 13(2):591–599, 2016.
- [143] A Al-Marakeby, Ayman A Aly, and Farhan A Salem. Fast quality inspection of food products using computer vision. *International Journal of Advanced Research in Computer and Communication Engineering* [1], 2, 2013.
- [144] GP Bu, S Chanda, H Guan, J Jo, M Blumenstein, and YC Loo. Crack detection using a texture analysis-based technique for visual bridge inspection. *Electronic Journal of Structural Engineering*, 14(1):41–48, 2015.
- [145] Zhiqiang Chen, RR Derakhshani, Ceki Halmen, and John T Kevern. A texture-based method for classifying cracked concrete surfaces from digital images using neural networks. In *Neural Networks (IJCNN), The 2011 International Joint Conference on*, pages 2632–2637. IEEE, 2011.
- [146] Vladimir P Vavilov and Douglas D Burleigh. Review of pulsed thermal ndt: Physical principles, theory and data processing. *Ndt & E International*, 73: 28–52, 2015.
- [147] Jeff R Brown and HR Hamilton. Quantitative infrared thermography inspection for frp applied to concrete using single pixel analysis. *Construction and Building Materials*, 38:1292–1302, 2013.
- [148] R Di Maio, C Mancini, C Meola, and E Piegari. Numerical modelling of architectonic structures thermal response. laboratory and in-situ data analysis. In *11th International Conference on Quantitative Infrared Thermography (QIRT)*, 2012.
- [149] Janet FC Sham, Tommy Y Lo, and Shazim Ali Memon. Verification and application of continuous surface temperature monitoring technique for investigation of nocturnal sensible heat release characteristics by building fabrics. *Energy and Buildings*, 53:108–116, 2012.

- [150] Frédéric Taillade, Marc Quiertant, Karim Benzarti, and Christophe Aubagnac. Shearography and pulsed stimulated infrared thermography applied to a nondestructive evaluation of frp strengthening systems bonded on concrete structures. *Construction and Building Materials*, 25(2):568–574, 2011.
- [151] DG Aggelis, EZ Kordatos, DV Soulioti, and TE Matikas. Combined use of thermography and ultrasound for the characterization of subsurface cracks in concrete. *Construction and Building Materials*, 24(10):1888–1897, 2010.
- [152] YY Hung, Yun Shen Chen, SP Ng, L Liu, YH Huang, BL Luk, RWL Ip, CML Wu, and PS Chung. Review and comparison of shearography and active thermography for nondestructive evaluation. *Materials Science and Engineering: R: Reports*, 64(5-6):73–112, 2009.
- [153] Ch Maierhofer, Ralf Arndt, Mathias Röllig, Carsten Rieck, Andrei Walther, Horst Scheel, and Bernd Hillemeier. Application of impulse-thermography for non-destructive assessment of concrete structures. *Cement and Concrete Composites*, 28(4):393–401, 2006.
- [154] Carosena Meola, Giovanni Maria Carlomagno, and Luca Giorleo. Geometrical limitations to detection of defects in composites by means of infrared thermography. *Journal of Nondestructive Evaluation*, 23(4):125–132, 2004.
- [155] A Wyckhuysen and Xavier Maldague. A study of wood inspection by infrared thermography, part i: Wood pole inspection by infrared thermography. *Research in nondestructive evaluation*, 13(1):1–12, 2001.
- [156] E Grinzato, V Vavilov, and T Kauppinen. Quantitative infrared thermography in buildings. *Energy and Buildings*, 29(1):1–9, 1998.
- [157] transformarobotics. Quicabot V4, 2016. URL <https://www.transformarobotics.com/>.
- [158] Jia Miin Yip, Naila Mouratova, Rebecca M Jeffery, Daisy E Veitch, Richard J Woodman, and Nicola R Dean. Accurate assessment of breast volume: a study comparing the volumetric gold standard (direct water displacement measurement of mastectomy specimen) with a 3d laser scanning technique. *Annals of plastic surgery*, 68(2):135–141, 2012.

- [159] David Fofi, Tadeusz Sliwa, and Yvon Voisin. A comparative survey on invisible structured light. In *Machine vision applications in industrial inspection XII*, volume 5303, pages 90–99. International Society for Optics and Photonics, 2004.
- [160] Viet Nguyen, Agostino Martinelli, Nicola Tomatis, and Roland Siegwart. A comparison of line extraction algorithms using 2d laser rangefinder for indoor mobile robotics. In *Intelligent Robots and Systems, 2005.(IROS 2005). 2005 IEEE/RSJ International Conference on*, pages 1929–1934. IEEE, 2005.
- [161] Bladimir Bacca, Joaquim Salvi, and Xavier Cufi. Appearance-based mapping and localization for mobile robots using a feature stability histogram. *Robotics and Autonomous Systems*, 59(10):840–857, 2011.
- [162] Elon Rimon and Daniel E Koditschek. Exact robot navigation using artificial potential functions. *IEEE Transactions on robotics and automation*, 8(5):501–518, 1992.
- [163] Ioannis Rekleitis, Vincent Lee-Shue, Ai Peng New, and Howie Choset. Limited communication, multi-robot team based coverage. In *Robotics and Automation, 2004. Proceedings. ICRA'04. 2004 IEEE International Conference on*, volume 4, pages 3462–3468. IEEE, 2004.
- [164] Howie Choset. Coverage for robotics—a survey of recent results. *Annals of mathematics and artificial intelligence*, 31(1-4):113–126, 2001.
- [165] Joe Sfeir, Maarouf Saad, and Hamadou Saliah-Hassane. An improved artificial potential field approach to real-time mobile robot path planning in an unknown environment. In *Robotic and Sensors Environments (ROSE), 2011 IEEE International Symposium on*, pages 208–213. IEEE, 2011.
- [166] Elena Garcia and P Gonzalez De Santos. Mobile-robot navigation with complete coverage of unstructured environments. *Robotics and autonomous systems*, 46(4):195–204, 2004.
- [167] Saroj Kumar Pradhan, Dayal Ramakrushna Parhi, and Anup Kumar Panda. Fuzzy logic techniques for navigation of several mobile robots. *Applied soft computing*, 9(1):290–304, 2009.

- [168] Mukesh Kumar Singh and Dayal R Parhi. Intelligent neuro-controller for navigation of mobile robot. In *Proceedings of the International conference on advances in computing, communication and control*, pages 123–128. ACM, 2009.
- [169] Auday Al-Mayyahi, William Wang, and Phil Birch. Adaptive neuro-fuzzy technique for autonomous ground vehicle navigation. *Robotics*, 3(4):349–370, 2014.
- [170] Petru Rusu, Emil M Petriu, Thomas E Whalen, Aurel Cornell, and Hans JW Spoelder. Behavior-based neuro-fuzzy controller for mobile robot navigation. *IEEE Transactions on Instrumentation and Measurement*, 52(4):1335–1340, 2003.
- [171] Hee Rak Beom and Hyung Suck Cho. A sensor-based navigation for a mobile robot using fuzzy logic and reinforcement learning. *IEEE transactions on Systems, Man, and Cybernetics*, 25(3):464–477, 1995.
- [172] Theodore W Manikas, Kaveh Ashenayi, and Roger L Wainwright. Genetic algorithms for autonomous robot navigation. *IEEE Instrumentation & Measurement Magazine*, 10(6), 2007.
- [173] Ashwin Ram, Gary Boone, Ronald Arkin, and Michael Pearce. Using genetic algorithms to learn reactive control parameters for autonomous robotic navigation. *Adaptive behavior*, 2(3):277–305, 1994.
- [174] Fredrik Gustafsson. Particle filter theory and practice with positioning applications. *IEEE Aerospace and Electronic Systems Magazine*, 25(7):53–82, 2010.
- [175] Gerasimos G Rigatos. Nonlinear kalman filters and particle filters for integrated navigation of unmanned aerial vehicles. *Robotics and Autonomous Systems*, 60(7):978–995, 2012.
- [176] Riccardo Poli. Analysis of the publications on the applications of particle swarm optimisation. *Journal of Artificial Evolution and Applications*, 2008, 2008.
- [177] Prases K Mohanty and Dayal R Parhi. Controlling the motion of an autonomous mobile robot using various techniques: a review. *Journal of Advance Mechanical Engineering*, 1(1):24–39, 2013.

- [178] Ren C Luo and Chih-Chia Chang. Multisensor fusion and integration: A review on approaches and its applications in mechatronics. *IEEE Transactions on Industrial Informatics*, 8(1):49–60, 2012.
- [179] Edouard Ivanjko, Mario Vasak, and Ivan Petrovic. Kalman filter theory based mobile robot pose tracking using occupancy grid maps. In *International Conference on Control and Automation (ICCA2005)*, volume 2, pages 869–874, 2005.
- [180] Jiali Shen and Huosheng Hu. Svm based slam algorithm for autonomous mobile robots. In *Mechatronics and Automation, 2007. ICMA 2007. International Conference on*, pages 337–342. IEEE, 2007.
- [181] Jingwen Tian, Meijuan Gao, and Erhong Lu. Dynamic collision avoidance path planning for mobile robot based on multi-sensor data fusion by support vector machine. In *Mechatronics and Automation, 2007. ICMA 2007. International Conference on*, pages 2779–2783. IEEE, 2007.
- [182] Dennis M Buede and Paul Girardi. A target identification comparison of bayesian and dempster-shafer multisensor fusion. *IEEE Transactions on Systems, Man, and Cybernetics-Part A: Systems and Humans*, 27(5):569–577, 1997.
- [183] Thierry Denoeux. A neural network classifier based on dempster-shafer theory. *IEEE Transactions on Systems, Man, and Cybernetics-Part A: Systems and Humans*, 30(2):131–150, 2000.
- [184] Yanmei Zhan, Henry Leung, Keun-Chang Kwak, and Hosub Yoon. Automated speaker recognition for home service robots using genetic algorithm and dempster-shafer fusion technique. *IEEE Transactions on Instrumentation and measurement*, 58(9):3058–3068, 2009.
- [185] Federico Castanedo. A review of data fusion techniques. *The Scientific World Journal*, 2013, 2013.
- [186] Bahador Khaleghi, Alaa Khamis, Fakhreddine O Karray, and Saiedeh N Razavi. Multisensor data fusion: A review of the state-of-the-art. *Information fusion*, 14(1):28–44, 2013.

- [187] Sayed Amir Hoseini and Mohammad Reza Ashraf. Computational complexity comparison of multi-sensor single target data fusion methods by matlab. *arXiv preprint arXiv:1307.3005*, 2013.
- [188] Xiangmao Chang, Rui Tan, Guoliang Xing, Zhaohui Yuan, Chenyang Lu, Yixin Chen, and Yixian Yang. Sensor placement algorithms for fusion-based surveillance networks. *IEEE Transactions on Parallel and Distributed Systems*, 22(8):1407–1414, 2011.
- [10] Erdal Kayacan I-Ming Chen Lee Kong Tiong Yan, Rui-Jun and Jing Wu. Quicabot: Quality inspection and assessment robot. *IEEE Transactions on Automation Science and Engineering*, 16(2):506–517, 2018.
- [189] J Heng. This building inspector misses nothing. <http://www.straitstimes.com/singapore/manpower/this-building-inspector-misses-nothing>, 2016.
- [190] Rui-Jun Yan, Erdal Kayacan, I-Ming Chen, and Lee Kong Tiong. A novel building post-construction quality assessment robot: Design and prototyping. In *2017 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pages 6020–6023. IEEE, 2017.
- [191] Suyog Dutt Jain and Kristen Grauman. Active image segmentation propagation. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 2864–2873, 2016.
- [192] ABS American Buireau of Shipping. Guidance notes on the inspection maintenance and application of marine coating system third edition. *Book*, 2007, 2007.
- [193] MSC Resolution. 215 (82),performance standard for protective coatings for dedicated seawater ballast tanks in all types of ships and double-side skin spaces of bulk carriers. *IMO, London, UK*, 2006.