Infrared machine vision and infrared thermography with deep learning: A review

Article *in* Infrared Physics & Technology · April 2021

DOI: 10.1016/j.infrared.2021.103754

7 authors, including:

Yunze he
Hunan University
104 PUBLICATIONS **2,954** CITATIONS
SEE PROFILE

Baoyuan Deng
Hunan University
6 PUBLICATIONS **26** CITATIONS
SEE PROFILE

Hongjin Wang
Hunan University
13 PUBLICATIONS **95** CITATIONS
SEE PROFILE

Francesco Ciampa
University of Bath
81 PUBLICATIONS **1,740** CITATIONS
SEE PROFILE

Infrared machine vision and infrared thermography with deep learning: a review

Yunze He[1], Baoyuan Deng[1,*], Hongjin Wang[1], Liang Cheng[2,3,*], Ke Zhou[1], Siyuan Cai[1], Francesco Ciampa[4]

1. College of Electrical and Information Engineering, Hunan University, Changsha, 410082, China;
2. School of Ocean Engineering, Jiangsu Ocean University, Lianyungang, 222005, China;
3. OceanAlpha Group LTD, Zhuhai, 519080, China;
4. Department of Mechanical Engineering Sciences, University of Surrey, Guildford, UK;

Abstract: Infrared imaging-based machine vision (IRMV) is the technology used to automatically inspect, detect, and analyse infrared images (or videos) obtained by recording the intensity of infrared light emitted or reflected by observed objects. Depending on whether controllable excitation is used during the imaging of infrared rays, thermal IRMV can be categorised into passive thermography and active thermography. Passive thermography is an important supplement to conventional machine vision based on visible light and is a valid imaging tool for self-heating objects such as the human body and electrical power devices. Active thermography is a non-destructive testing method for the quality evaluation and safety assurance of non-self-heating objects. In active thermography, the trend is to inspect rapidly, reliably, and intelligently by introducing multiple-mode excitation sources and artificial intelligence. The rapid development of deep learning makes IRMV more and more intelligent and highly automated, thus considerably increasing its range of applications. This paper reviews the principle, cameras, and thermal data of IRMV and discusses the applications of deep learning applied to IRMV. Case studies of IRMV and deep learning on various platforms such as unmanned vehicles, mobile phones and embedded systems are also reported.

Keywords: Machine vision, Deep learning, Thermography non-destructive testing (TNDT), Unmanned aerial vehicle (UAV), Object detection, Semantic Segmentation

Appendix: Abbreviation

AE	Autoencoder
AT	Active thermography
CFRP	Carbon Fiber Reinforced Polymer

* Corresponding author.
E-mail address: dengbaoyuan@hnu.edu.cn (B. Deng)

CNN	Convolutional Neural Network
DAN	Deep alignment network
DINN	Deep inception neural network
DL	Deep learning
DNNs	Deep neural networks
EC	Edge computing
FIR	Far infrared
GAN	Generative adversarial network
GFRP	Glass Fiber Reinforced Polymer
IoU	Intersection over union
IRMV	Infrared machine vision
IRT	Infrared thermography
LST	Line scanning thermography
LSTM	Long short-term memory
LT	Lock-in thermography
LWIR	Long-wavelength infrared\
mAP	Mean average precision
MAV	Micro-aerial vehicle
ML	Machine learning
MLP	Multilayer perceptron
MUT	Material under test
MT	Modulated thermography
MV	Machine vision
MWIR	Mid-wavelength infrared
NDT	Nondestructive testing
NIR	Near-infrared
OLT	Optical lock-in thermography
PPT	Pulsed phase thermography
PSNR	Peak signal-to-noise ratio
PT	Pulsed thermography
RBM	Restricted Boltzmann machine
RNN	Recurrent neural network
RIRMV	Reflected infrared machine vision
SRCNN	Super-resolution convolutional neural network
SSD	Single shot detection
SSIM	Structural similarity index measure
ST	Stepped thermography
SWIR	Short-wavelength infrared

TIRMV	Thermal infrared machine vision
UAV	Unmanned aerial vehicles
VL	Visible light

1. Introduction

Human eyes are sensitive to electromagnetic rays with wavelengths ranging from 400 to 760 nm (i.e. the visible light, VL). Machine vision (MV) is the technology that enables imaging machines to inspect, recognize and analyse images as human beings. A typical MV-based system is shown in Fig.1(a), which includes VL cameras with lenses, light sources, personal computers (PCs, or embedded system) for image processing and actuators. Couple charged devices (CCDs) and complementary metal oxide semiconductors (CMOSs) are prevalent technology for capturing images, from digital astrophotography to machine vision inspection. Silicon based CCDs or CMOSs can be sensitive to a range of wavelengths as in human's eye, but a litter wider. However, there is still a wide range of electromagnetic waves that cannot be sensed with CCD or CMOS based cameras. Thus, some MV systems expand the functions at infrared (IR), ultraviolet (UV), or X-ray wavelengths by using different photon-electrical materials or even thermal-electrical materials [1, 2]. Hence, infrared machine vision (IRMV) is an important supplement to MV. Devices included in a typical IRMV system are infrared cameras and lenses sensing the IR rays from 760 nm to 1 mm in wavelength, PCs for image processing and controllers. Excitation sources, instead, are optional and should be used depending on the actual application. Such distinction of thermal IRMV systems led to both passive thermography (PT) and active thermography (AT). PT can be used to obtain thermal images of an object that is constantly and naturally in disequilibrium from its environment without a controllable excitation source. AT, instead, is a well-established non-destructive testing (NDT) method that is used when internal defects need to be assessed in thermal-equilibria materials by exciting with an external thermal source to create a thermal gradient into the material.

It can be seen that both conventional MV based on visible light and IRMV based on IR rays are inseparable with machine learning (ML). In recent years, as for the rapid development of deep learning (DL), particularly with a class of ML algorithms that use multiple layer networks to progressively extract higher level features from the raw input, both MV and IRMV have obtained increasing interest in the fields of medicine, industry and architecture [3-6]. The rapid development of deep learning makes IRMV more intelligent and automated thus enhancing its range of applications. After reviewing the related papers published in the recent 3-5 years, this work introduces firstly and systematically the principle of IRMV, cameras, excitation sources and data processes used in IRMV. In Section 2, DL models in IRMV are reviewed. Then, the applications of DL in passive TIRMV in different areas are presented and compared in Section 3. The application of DL in active TIRMV for thermography NDT is reported in Section 4. The DL based on various platforms such as unmanned aerial vehicles (UAV), mobile phones and embedded systems are discussed in Section 5. Section 6 forecasts the

development trends of IRMV. Finally, conclusions are drawn in the last Section.

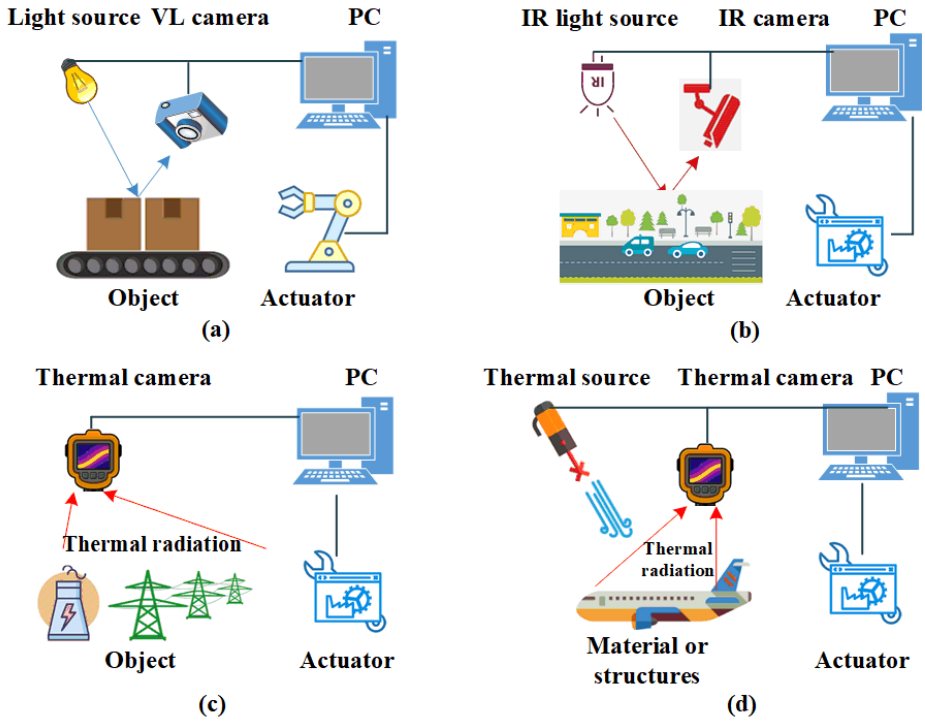

Fig. 1. Diagram of MV and IRMV. (a) conventional MV, (b) reflected infrared machine vision, (c) passive thermal infrared machine vision, and (d) active thermal infrared machine vision

2. Principle, Camera, Sources and Data of IRMV

Infrared-imaging based machine vision (IRMV) is the technology used to automatically inspect, detect, and analyse infrared images (or infrared videos) obtained by recording the intensity of IR light emitted or reflected by observed objects. An alternative definition of IRMV is the ability of machines to create images by infrared (IR) rays radiated or reflected by objects, where IR rays are electromagnetic waves with wavelengths ranging from 760 nm to 1 mm. Through the information extracted from these images, IRMV makes decisions and controls the behaviour of a machine systems automatically. As shown in Fig. 1(b), a typical IRMV system consists of an IR camera with lenses, IR light sources, a PC for image processing and control, and actuators. There is a clear distinction between IRMV and MV in terms of physical principle, cameras, and light sources.

2.1 Principle of IRMV

Compared to visible light, infrared rays spread through a board wavelengths range. Thus, the infrared band is often subdivided into subsections based on their wavelength. Usually, near-infrared (NIR) rays are referred to the electromagnetic waves with wavelength ranges from 0.76 μm to 1 μm, short-wavelength infrared (SWIR) from 1 μm to 3 μm, mid-wavelength infrared (MWIR) from 3μm to 8μm, long-wavelength

infrared (LWIR) from 8μm to 15μm and far infrared (FIR) from 15 to 1000 μm. Hence, there are five different classifications of MV according to the sub-division scheme, as shown in Table 1. FIR rays, in particular, with wavelengths ranging from 0.1 mm to 1 mm are called terahertz (THz) rays. Thus, terahertz machine vision (THzMV) could be defined based on THz camera like TeraSense [7, 8].

Table 1. The principle and classification of IRMV.

Kinds	**Subclass**	**IR rays**	**Wavelength (μm)**	**Common sensor**	**Sources**
RIRMV	NIRMV	NIR	0.76-1	Ge/Si CCD, COMS	LED, Laser
	SWIRMV	SWIR	1-3	InGaAs	LED, Laser
TIRMV	MWIRMV	MWIR	3-8	InSb, MCT, QWIP	Not required in passive mode; Flash, lamp, laser, eddy current, microwave, ultrasound in active mode
	LWIRMV	LWIR	8-15	Microbolometer	
RIRMV	FIRMV	FIR	15-1000	Microbolometer, ZnTe, GaAs, GaN, etc	FIR light
	THzMV	THz	100-1000	ZnTe, GaAs, GaN, etc.	THz rays

It is well known that, due to thermal radiation, all objects in the universe with temperature above the absolute zero Kalvin degree can emit electromagnetic waves in an amount correlating to its surface temperature. In specific, the electromagnetic waves radiated from an object (normally -40℃~400℃) on earth are mainly in the range of MWIR and LWIR, which cause significant heating effect to the objects in the nature earth environment. Thus, MWIR and LWIR is sometimes referred to as "thermal infrared" while NIR and SWIR is sometimes called "reflected infrared" since the major part of the latter waves are reflected by most of objects in the nature earth. Infrared machine vision can be classified into thermal infrared machine vision (TIRMV) and reflected infrared machine vision (RIRMV). In the restrict but widely used definitions, TIRMV includes just MIRMV and LIRMV while RIRMV includes NIRMV, SWIRMV and FIRMV. It appears that such a category of IRMV technology is divided based on the wavelength of infrared rays which the system senses, but in fact such a category is divided whether the decision made by an IRMV system relies on thermal information underlying the IR images. As a result, in a broad definition, SWIRMV, FIRMV and THzMV can also be classified as TIRMV when the technology is used to identify the surface temperature of the objects.

TIRMV is the focus of this review, which is also known as infrared thermography (IRT) or infrared thermal imaging, because the decision maker in such a system highly relies on the thermal information behind the infrared-rays formed images rather than

the insensitivity of the infrared rays itself. As reported above, TIRMV can be classified into passive TIRMV (as shown in Fig.1(c), or passive thermography) and active TIRMV (as shown in Fig. 1(d), or active thermography) according to that whether a controllable excitation source is required. The key point of this work is how to apply DL to post data processing of active thermography. Thus, the principles of these AT methods are slightly introduced and more detailed knowledge can be found in previous work [9].

2.2 Camera of IRMV

IR cameras and lenses are necessary in any kind of IRMV. As shown in Table 1, the common camera used in NIRMV is Ge/Si based CCD /CMOS cameras, because they can sense not only VL but also the IR rays from 0.76 to 1.1 um with a weak sensitivity. The common cameras used in SWIRMV are InGaAs based camera, which is sensitive to IR rays from 0.9 to 1.7 um. The commonly used cameras in MWIRMV are made of InSb, MCT and QWIP, which need cooled environment and is sensitive to IR rays from 3 to 5 um (with the ~20mK in NETD). The common cameras used in LWIRMV are made of VO_x or SiO_x based microbolometer or micro-pyramids, which can be sensitive to IR rays from 8 to 14 um (with the ~50mK in NETD) without cooling. The electrical signal in the pixel of an uncooled microbolometer detector decays exponentially with an integral period of 10–15ms [10]. The thermal inertia leads to microbolometer having NETD that were inferior by factors of between approximately 2 and 6 [11]. Higher NETD also means higher price. In the application where quick acquisition is need, thermal inertia aggravates the motion blur when there is relative motion using an uncooled camera [12], while motion blur is not obvious using a cooled camera [13].

For simplification, these cameras are all referred as IR cameras. The cameras in TIRMV is also named thermal camera (thermal imager) because they measure the temperature of the observed object and form images based on thermal information. To iterate, this work mainly reviews on narrowly defined TIRMV (MWIRMV and LWIRMV). Currently, thermal cameras have some disadvantages compared to VL camera, such as low resolution (normally <640×512), low frame rate (about hundreds Hz) and high cost (thermal camera is still more expensive than VL camera having the same resolution). Even though, thermal IRMV has widely investigated in many fields. In recent years, some portable thermal cameras based on mobile phone and embedded system appears, which promote the development of TIRMV[7].

2.3 Sources of IRMV

Excitation sources are generally required in IRMV systems since the observed objects are in a thermal equilibrium state with their environment at the time of the inspection. As shown in Table 1, the common sources used in NIRMV and SWIRMV is light sources like LED or laser which can emit IR rays from 0.76 to 3 um. The common used sources in the active mode of TIRMV are thermal sources like flash lamps and halogen lamps etc. [9]. However, in the passive mode of TIRMV, sources

are unnecessary because the object can radiate the thermal infrared rays by itself. The common sources used in FIRMV and THzMV are radiation sources which can emit FIR lights and THz rays respectively.

There many thermal sources with different shapes that can be selected in AT. The developing trends in AT is to apply the multi-mode excitation, time domain modulation and spatial modulation, as shown in Fig. 2. According to the nature of thermal sources, active thermography can be classified into flash thermography, lamp thermography, laser thermography, eddy current thermography, microwave thermography and vibro-thermography etc. [9, 14, 15]. The mechanisms behind these active thermography methods are different from each other. For example, eddy current thermography, microwave thermography and vibro-thermography employs volumetric heating or inside heating [16] as excitation sources.

AT can also be classified into pulse thermography, step thermography, long-pulse thermography [17-19], lock-in thermography and frequency modulation thermography according to the turn-on scheme of the excitation sources. Different heating methods correspond to different post-processing methods. For example, the heating data is valued in step thermography, while pulse thermography and long-pulse thermography refer to the cooling stage [18]. In future, thermal sources can be modulated in the space domain to form square-like, point-like, line-like and array heating areas [20].

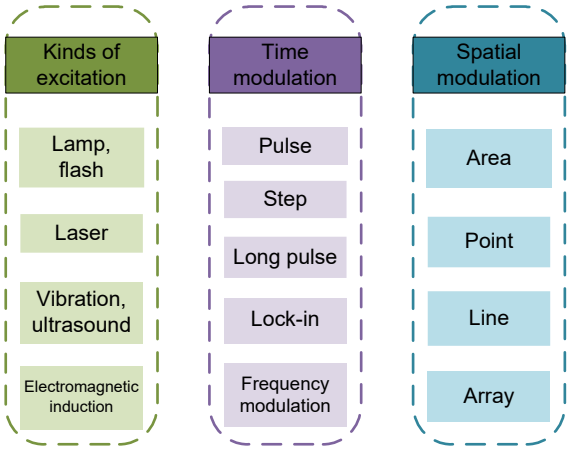

Fig. 2. Typical excitation sources used in AT.

2.4 Data of IRMV

In both passive and active TIRMV, the data captured by thermal cameras are passed to the image processing system in the form of thermal image sequences or thermal videos. As shown in Fig. 3, there are three deep learning models: (1) 1D model [such as 1D Convolutional Neural Network (CNN) and Recurrent Neural Network (RNN)], where the input is the vector of temperature series regardless the spatial information contained by the thermal images or videos; (2) 2D model, where the input is the thermal images and the common model is 2D CNN; (3) 3D model, where the input is the video or thermal image sequence and the common model is 3D CNN, whose application will be reviewed in Sections 3 and 4.

The number of labelled IR images is much smaller compared to the visible images.

It is a very interesting task to reduce the amount of data required for IRMV using labelled VL images. Transfer learning [21, 22] is a research problem in machine learning (ML) that focuses on storing knowledge gained while solving one problem and applying it to a different but related problem. If only the data are different and the tasks are the same, domain adaptation [23-25] is a very efficient way of transfer learning. For example, Popular VL tasks such as object detection [25, 26], face recognition[27-29] etc. provide very valuable experience and data for passive IRMV.

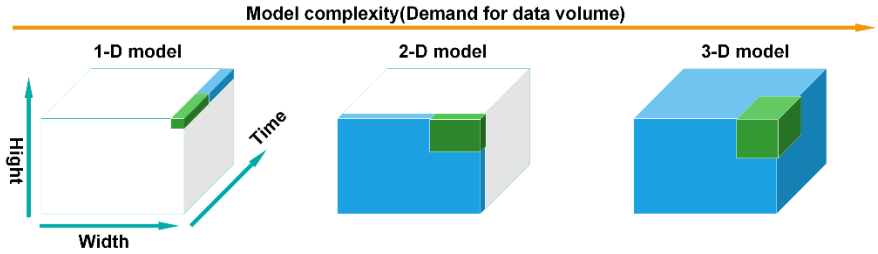

Fig. 3 Typical data of the IRMV method

3. Deep learning in passive thermography

The usage of DL models including CNN, RNN, Autoencoder (AE), Restricted Boltzmann Machine (RBM) and Generative Adversarial Network (GAN) in PT are quite more often than those in AT. Typical cases are introduced and analysed according to the specific application.

3.1 Image super-resolution and quality improvement

As mentioned in Section 2.1, the resolution of thermal cameras is lower than VIC (millions of pixels), which leads to a poor-detailed image and a limited application of PT. Some researchers have investigated CNN [30, 31] and GAN[32, 33] to improve the quality of thermal images [34].

Some scholars consider taking single frame super-resolution technology to improve the resolution of IRT images. In Ref. [30], a method of quality improvement in IRT images for object recognition was investigated. A systematic approach based on image bias correction and super-resolution CNN (SRCNN) was proposed to increase target signature resolution and optimize the baseline quality of inputs for the object recognition. The main objective is to maximize the useful information on the object to be detected even when the number of target pixels is adversely small. Experimental results showed that the approach can significantly improve target resolution and thus helps making object recognition more efficient in automatic target detection/recognition systems.

In particular, using super-resolution technique to improve the resolution facial area images is proposed in Ref. [31]. Several network structure has been tested, and Fig. 4 demonstrated the deep residual embedding and supervised-recursion (DRESNet) , which has the highest peak signal-to-noise ratio (PSMR). Besides the attempt in network structure, this work also emphasizes the importance of high accuracy data. The

performed evaluation showed that PSNR can be improved even by 60% if 16 bit depth resolution data is used instead of 8 bits.

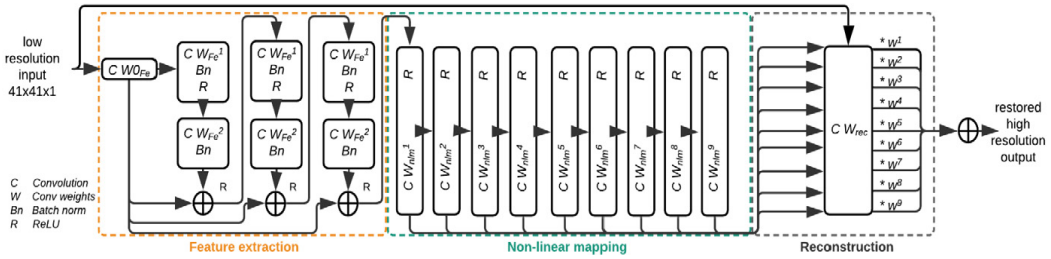

Fig. 4. The best configuration of the DRESNet in Ref. [31]

Rather than simply increasing the resolution of IRT, some scholars try to utilize the complementarity of VL and IRT to fuse the information of the two for super-resolution. In Ref. [32], a deep learning solution based on GAN was created to enhance the thermal image resolution with the fusion of VL and IRT images. Result showed that the quality of visual-thermal fusion can be enhanced by the GAN based model. Besides, this paper indicate the shortcoming of widely used PSMR measurement by a qualitative evaluation.

In order to enhance the visualization of infrared and visible image fusion, a method combining a generative adversarial network (GAN) and a residual network (ResNet) is proposed [33]. The illustration of this work is shown in Fig. 5. The concatenation of infrared and visible image is regarded as the input the generator network. Experiment results demonstrate that the proposed method eventually gets desirable images, achieving better performance in objective assessment and visual quality compared with nine representative infrared and visible image fusion methods.

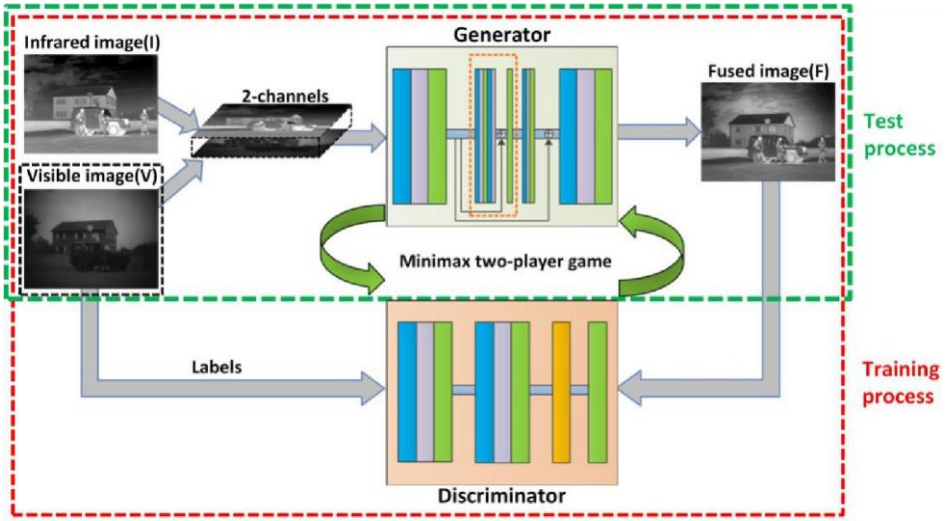

Fig. 5. The framework of infrared and visible image fusion [33]

3.2 Medical Research

IRT with MV is widely studied to detect breast cancer and diabetic foot in a non-invasive and non-contact way. In study [35], the detection of breast cancer is regarded as a classification problem. A LeNet-based CNN is utilized to classify automatically

and results from this study indicated that CNNs obtained competitive results for both static and dynamic protocols.

Some other scholars regard the detection of breast cancer in IRT images as a segmentation problem. In Ref. [36], autoencoder-like convolutional and deconvolutional neural networks (C-DCNN) are promising computational approaches for automatic breast areas segment in IRT images, which will help circle out the area for tumour search and reduce the time and effort needed for manual segmentation. Through cross-validation and comparison with the ground-truth images, results demonstrate the capability of C-DCNN to learn essential features of breast regions and delineate them in thermal images, showing that the C-DCNN is a promising method to segment breast regions.

Considering that the captured IRT images may vary in viewpoint, a cascaded CNN architecture is proposed to perform accurate segmentation in Ref. [37]. Through the data augment in images capturing stage, this method is robust to subject views and capture errors. The evaluation indicates that this approach can detect breasts region independently of the image capture and view angle, thus enabling automated image and video analysis.

In the early inspection of diabetic foot, patch-wise CNN [38] and R-CNN [39] has been attempted. In Ref. [38], the complete diabetic foot image is divided into several patches as the input the network due to the small dataset of 110 images. Machine learning-based techniques support vector machine (SVM) and multilayer perceptron (MLP) is compared with DL structures AlexNet, GoogLeNet, and the proposed CNN. The proposed Diabetic Foot Thermograms Network (DFTNet) has higher AUC-values of 0.8333, showing the usefulness to aid during the medical diagnosis.

In order to enhance the adaptability of the segmentation algorithm, two datasets is established in Ref. [39]. DB1 captured by FLUKE TI32 IRT camera and DB2 captured by mobile phone camera. The R-CNN model is trained with DB2 and validated by DB1 for the purpose of foot sole segmentation. The results illustrate a detection accuracy of 90% for ulcers and 88% for necrosis respectively.

3.3 Object detection and tracking

Detecting and tracking objects are among the most prevalent and challenging tasks that a surveillance system has to accomplish in order to determine meaningful events and suspicious activities, and automatically annotate and retrieve video content [40]. In this section, we major to investigate three types of tasks among object detection and tracking.

For the task of segmentation of background and objects, semantic segmentation networks U-Net and V-Net is applied for the segmentation of thermal images of laboratory rats that are in close contact during social behaviour tests in Ref. [41]. This study emphasizes the effects of selected temperature range of input image. Through lots of experiments and evaluations, they find out that the raw image data does not perform best compared to images with incomplete temperature ranges and the different model fits different ranges even for the same task. By combining VL and IRT images, a semantic segmentation called Multi-spectral Fusion Networks (MFNet) is proposed in

Ref. [42] for the semantic segment of images of street scenes for autonomous vehicles based on a new RGB-Thermal dataset. The original intention of MFNet is to achieve a good balance between accuracy and inference speed. Results revealed that MFNet can be small and sufficiently fast to achieve real-time performance 55 images/s on NVIDIA Geforce Titan X GPU. At the same time, similar or higher accuracy than state-of-the-art segmentation methods such as SegNet was achieved. This work combining RBG and thermal imagery provided a solution of poor visibility at night and under adverse weather conditions.

Salient object detection is a task based on a visual attention mechanism, in which algorithms aim to explore objects or regions more attentive than the surrounding areas on the scene or images. In Ref. [43], a novel end-to-end network for multimodal salient object detection is proposed, which turns the challenge of RGB-T saliency detection to a CNN feature fusion problem. Using collaborative graph learning approach for RGB-T image saliency detection has been proposed in Ref. [44]. In particular, super pixels are taken as graph nodes, hierarchical deep features to jointly learn graph affinity and node saliency in a unified optimization framework are used collaboratively. Experiments on several datasets have demonstrated the effectiveness of these two approaches.

For the task of single-frame template matching and consecutive object tracking, CNN-based approach has also been adopted. Single frame template matching using VGG feature has shown great improvement on compared to the histogram of gradients (HOG) feature [45].

Demonstrated in Fig. 6, a tracking method based on dynamic Siamese network and the fusion of visual and thermal image feature is proposed [46]. Visible and infrared images are firstly processed by two dynamic Siamese Networks, namely visible and infrared network, respectively. Then, multi-layer feature fusion is performed to adaptively integrate multi-level deep features between visible and infrared networks. Response maps produced from different fused layer features are then combined through an elementwise fusion approach to produce the final response map, based on which the target can be located. The results show very competitive performance against the state-of-art RGB-T trackers.

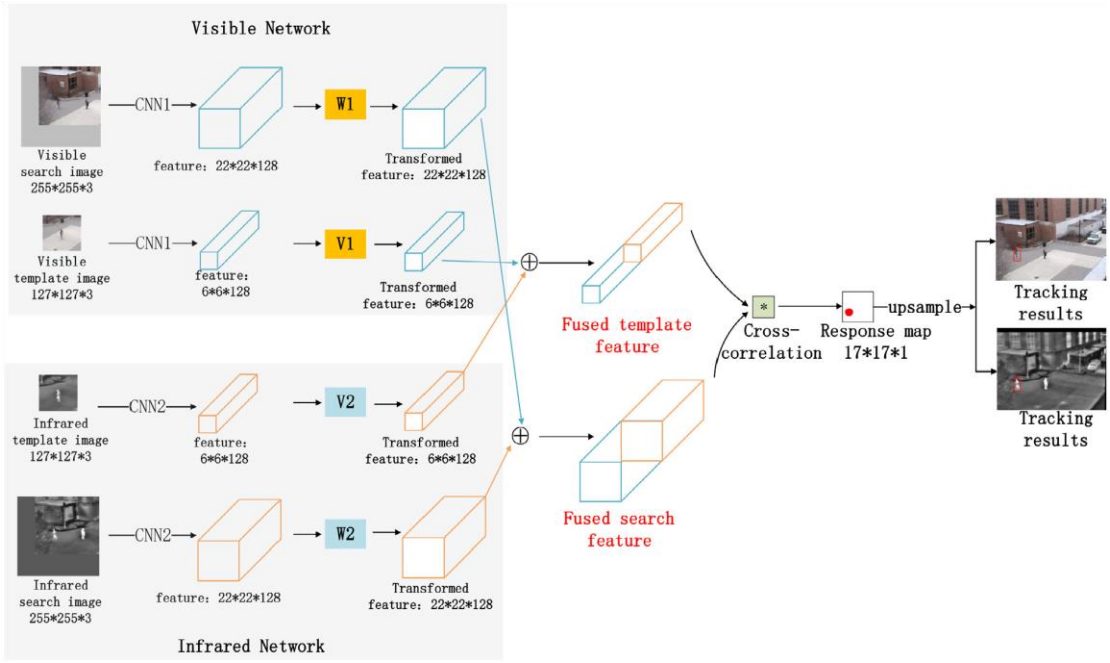

Fig. 6. Flowchart of the fusion tracking algorithms based on dynamic Siamese networks in Ref. [46]

3.4 Pedestrian detection and tracking

Pedestrian detection and tracking are essential tasks for assisted driving and autonomous driving. Because of the characteristics of IRT, IRT shows the advantages over VL in outdoor uncontrolled environment, such as night-time driving and bad illumination conditions [47].

In Ref. [26], CNN-based object detector frameworks, the Single Shot MultiBox Detector (SSD) with VGG16 as backbone, which are pre-trained on visual RGB images, was used to detect pedestrians in thermal imagery. The training strategy shown in Fig. 7, used transfer learning, where two key steps were undertaken. Firstly, an appropriate pre-processing strategy for the IR data was suggested, which transformed the IR data as close as possible to the RGB domain. This allowed pre-trained RGB features to be effective on the novel domain. Then, the remaining domain gap was addressed by fine-tuning the pre-trained CNN on a limited set of thermal IR data. Experiments indicated significant person detection improvements on the public KAIST dataset with the optimized pre-processing strategy. The results indicated that the domain differences between RGB and thermal IR data can be mitigated well by the small number of proposed steps based on RGB-pretrained detector networks.

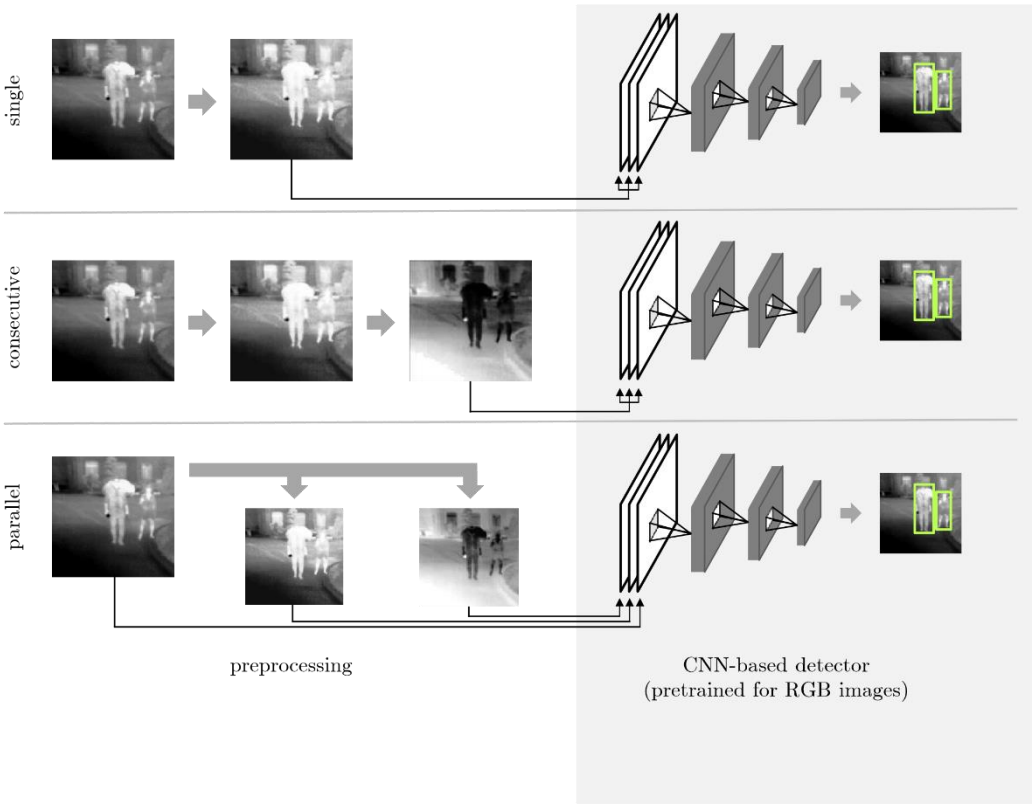

Fig. 7. Comparison of pre-processing methods in Ref. [26]. Top: no pre-processing concatenation, thermal image duplicates twice as input channels of the RGB pretrained detector; Mid: consecutive application of two pre-processing method; Bottom: three parallel pre-processing methods and concatenation.

In Ref. [48], YOLOv3 is used to recognize the people playing soccer or performing related exercises in outdoor sports fields. The data collection procedure is designed elaborately considering different viewpoints, indoor scene, and weather affects, taking 20 months. Through comparison test, model trained with the most data has the highest F1 score. This result highlights the importance of dataset variety in transfer learning. The pedestrian detection in severe weather conditions has been studied in Ref. [47] using YOLO. Besides the importance of dataset, Ref. [49] highlight the importance of mining auxiliary information between modes in cross-modal pedestrian retrieval. Cross-modal feature distribution and contextual information is proposed to reduce the cross-modal discrepancy and improve the detection results. The experiment results on several cross-modal dataset has shown its effectiveness.

3.5 Face recognition

Face recognition is a vital important task in automatic identity recognition, mobile payment, security protection. For thermal facial recognition, a thermal face detection and tracking method is proposed using Inception v3 network n Ref. [50]. This study majors to tackle the data from embedded thermal camera FLIR Lepton with a resolution of 80×60. Through the detection of nostril area using image patches, this method has achieved the face tracking task in a very short time of 42ms.

Along with face detection, another important topic is face alignment using landmark.

In [51], deep alignment network (DAN), a multistage approach in which several stages refine landmark positions predicted by the previous stage, is used to detect a set of landmarks of thermal infrared face. Evaluation has shown that the learning-based approaches, several of which have not been used for these tasks in the thermal infrared domain before, clearly outperform previously presented methods. The author concluded that a sufficiently large and well annotated database can be used to train different learning-based algorithms, which should be preferred over algorithm-based approaches due to their increased performance.

Besides thermal facial recognition, cross-modal face recognition has attracted more attention of researchers [27-29]. A deep coupling learning architecture with two-stream CNN is proposed in Ref. [27], whose network structure is shown in Fig. 8. Through large training data and deep network, it has achieved better performance than the swallow ones.

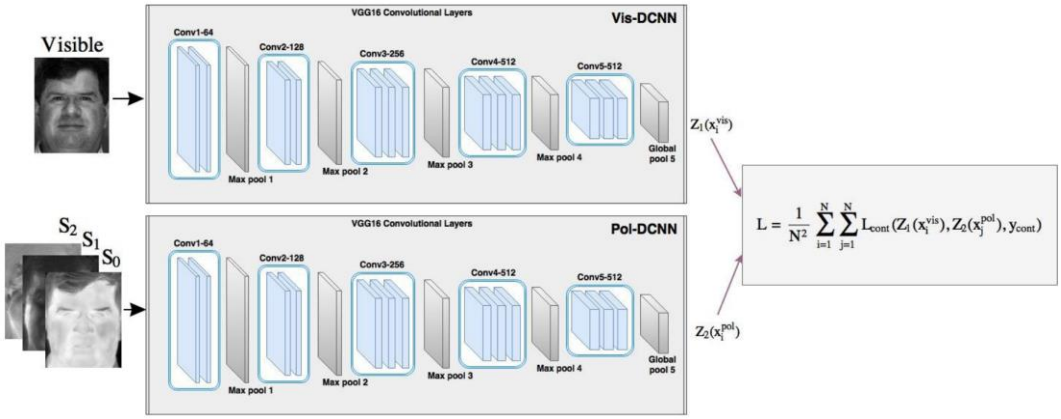

Fig. 8. Two-stream CNN in Ref. [27] for cross-modal face identification

A GAN based multi-stream feature-level fusion technique to synthesize high-quality visible images from polarimetric thermal images is proposed in Ref. [28]. The architecture of network in illustrated in Fig. 9. The generator contains a multi-stream feature-level fusion encoder–decoder network. In addition, a deep-guided subnet is stacked at the end of the encoding part. The discriminator is composed of a multi-scale patch-discriminator structure. This idea of generation inspires that using the generated VL face image to compare with the real VL face image for cross-modal face recognition. In Ref. [28], using generated VL face image for cross-modal face identification Is proposed and shown in Fig. 10. A GAN is trained generate VL face images from IRT images, and then two-stream CNN like Ref. [27] is used to determine.

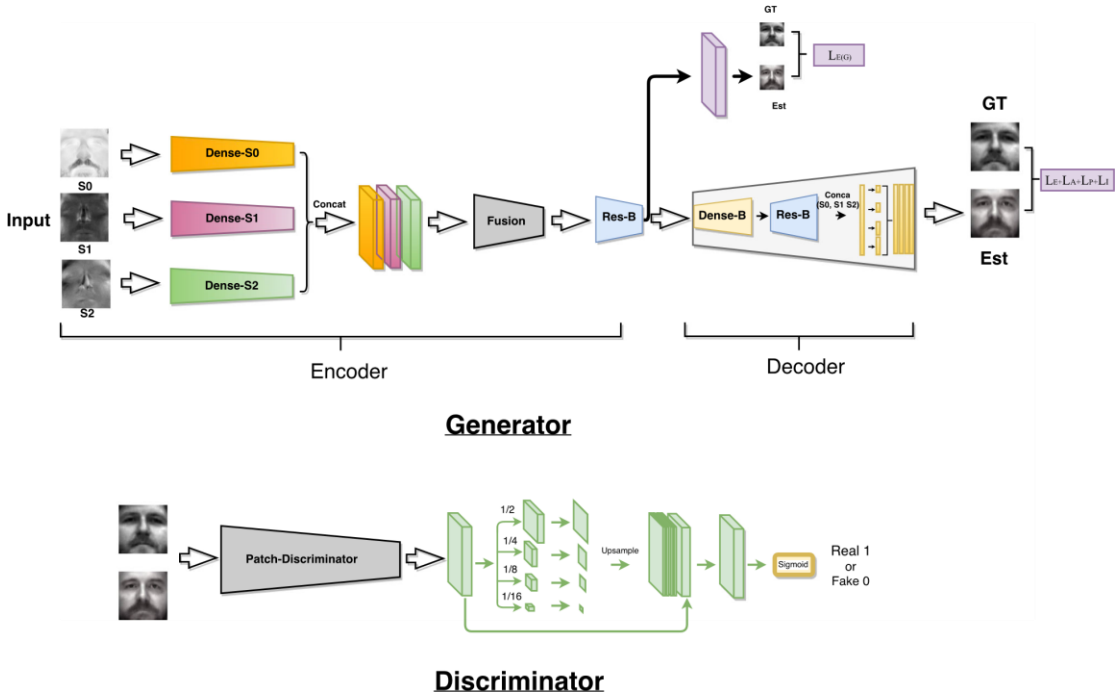

Fig. 9. structure of GAN in Ref. [28] for VL face image generation

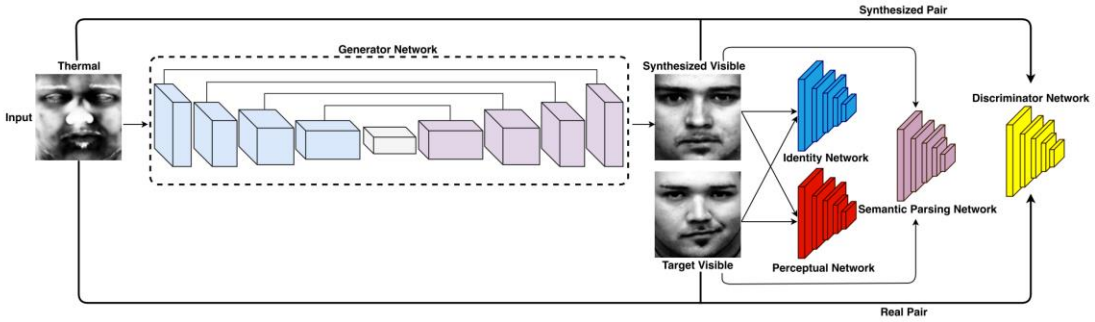

Fig. 10. The workflow of Ref. [29] for cross-modal face recognition

3.6 Behaviour detection

In [52], deep Learning-based action recognition method that combines convolution layers and an LSTM layer is proposed to learn the spatio-temporal representation of human actions whose inputs are thermal images and their frame differences cropped by the gravity centre of human regions. In extremely low-resolution images, the proposed method has achieved 91.07% accuracy of action classification overall.

Sign language is a visual language used by persons with hearing and speech impairment to communicate through finger spellings and body gestures. A framework for automatic sign language recognition is proposed without the need of hand segmentation [53]. Ensemble learning schemed in Fig. 11 is introduced with fine-tune three CNNs trained on ImageNet dataset. MobileNet is used as the backbone for feature extraction and kernel-based extreme learning machine (KELM) is proposed for feature classification and fusion. The experimental results show that the proposed ensemble learning is robust for human motion recognition on both visual and thermal data.

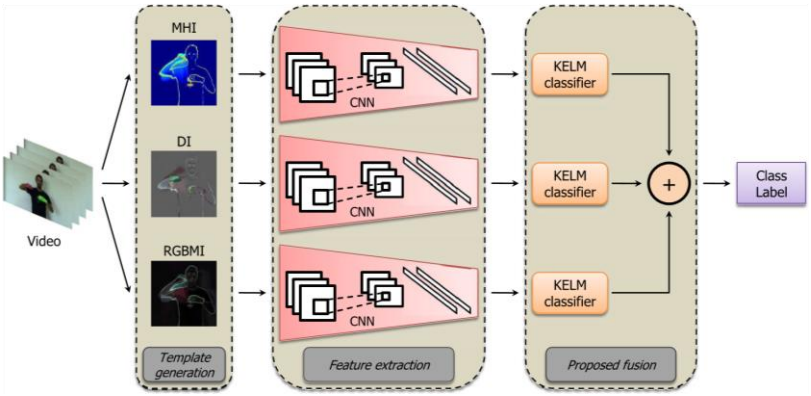

Fig. 11.Proposed ensemble model in Ref. [53] for sign language classification

3.7 Electrical equipment and machine

In [54], a deep CNN method that predicts the coordinates, orientation angle, and class type of each equipment part was proposed. A prior concerning orientation consistency between parts is also incorporated into the proposed model to improve the prediction results. For the evaluation, a large image set containing various types of scenarios was constructed. Experiments on the dataset showed that the method is robust to noise, achieving 93.7% mean average precision when the intersection over union threshold was 0.5, and running at 20 fps on GPU. The author believe that the high accurate detection results can benefit the subsequent diagnosis.

Early prevention measures of thermal anomalies of high voltage electrical equipment is essential for grid safety. A deep learning image-based model of integrating CNN and DL is proposed to determine whether the equipment is defective[55]. In this study, rich features are extracted from the convolutional layer of pretrained AlexNet and these feature vectors are classified using random forest. The experimental results with 2000 images outperforms the other methods, such as LeNet and VGG, reaching an accuracy of 96%.

3.8 Structural Health Monitoring

In [56], a new deep learning-based method is proposed to detect subsurface damage of steel members in a steel truss bridge using infrared thermography (IRT). With a LWIR FLIR T650SC, a total of 35 different thermal videos and several thermal images were collected and converted to approximately 2000 images with a resolution of 299×299 pixels and 200 images with a resolution of 640×480 pixels. To reduce computation costs, the original deep inception neural network (DINN) is modified for transfer learning. The proposed method provides bounding boxes for detecting and localizing subsurface damage such as corrosion and debonding between paint with coating and steel surface. Robustness and accuracy were tested on 200 thermal images, and 96% accuracy and 97.79% specificity were achieved.

The combination of thermography and CNN has also applied for rotating machinery fault diagnosis in Ref. [57]. The LeNet-like CNN is designed to classify the health

condition. The results validated by two different experimental data reveal that the proposed method has a superior performance in identification various faults on rotor and bearings comparing with other deep learning models and traditional vibration-based method.

3.9 Remote sensing

In [58], the CNN-based feature-learning strategy is proposed for the fusion of hyperspectral thermal infrared (HTIR) and visible remote sensing data. Experimental results showed that, except for the computational time, the proposed deep learning model outperformed shallow feature-based strategies in the classification performance that was based on its accuracy.

3.10 Photovoltaic inspection

In [59], deep learning and feature-based approach for the purpose of detecting and classify defect photovoltaics modules is examined in outdoor power plant of South Africa. The results show that the VGG-16 and MobileNet models have better performance compared to the combination of scale invariable feature transform (SIFT) descriptor and random forest classifier.

In [60], transfer learning is used with pre-trained VGG-16 model from electroluminescence images dataset of photovoltaic cells. The results show that the transfer learning technique can help to improve the performance to achieve an average accuracy of 99.23%.

3.11 Quality Control

In Ref. [61], the combination of IRT and CNN is used for the quality inspection of honey adulterations. The cooling process of 56 samples of pure honey have been mixed with different concentrations of rice syrup samples is monitored by IRT camera Optris PI 450 and the resulting images are subjected to a VGG-like CNN model. The model is trained and validated by the captured thermograms. The resulting model is capable of identifying adulterated honey from different floral origins and quantifying rice syrup with accuracies of 95% and 93%, respectively. The same method has been applied to inspect the adulteration state of extra virgin olive oil [62], reaching an accuracy range from 97 to 100%. The combination of thermography and CNN has also been used for the detection and classification of bruises of pears, reaching a best accuracy of 99.25% [63].

3.12 Materials

Inverse problems involving heat conduction are ubiquitous in engineering practice, but their solution is often challenging. A data-driven approach for the segmentation of heterogeneous composites is proposed [64]. The semantic segment model U-Net shown in Fig. 12 is used to classify the image into defect and non-defect pixel-wisely. 1200

datasets of temperature fields have been generated using simulation. When the true temperature at each pixel is given, the trained model can predict the distribution of fillers with an average accuracy of over 0.979, which suggests the feasibility of solving inverse problems with high-dimensional input and output using deep learning models.

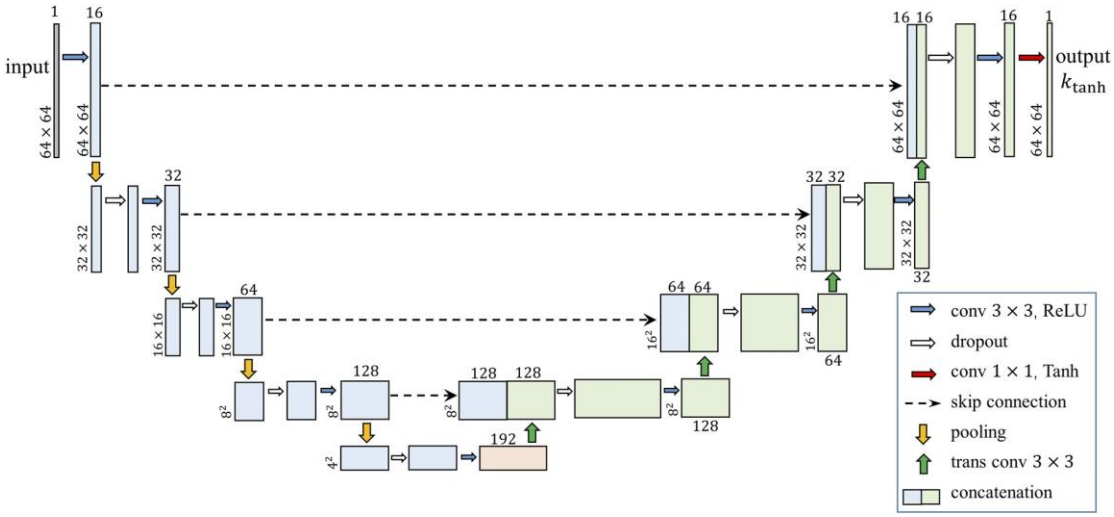

Fig. 12. The U-Net architecture in Ref. [64]. Blue boxes correspond to the feature map for the encoder, and the green boxes correspond to the feature maps for the decoder. The orange box denotes the bottleneck. The number of channels is marked on the top of each box; the x–y size (weight and height of the feature map) is provided at the lower-left edge of the box. The arrows denote the different operations explained in the legend.

Table 2. the specific parameters of DL in passive TIRMV

Application	Reference	Objects	Datasets	Input	Task	DL model	Model	Highlight	Metrics
Image super-resolution and quality improvement	Zhang [30], 2016	Car	(2197+575) images; 640×512	32×32	image quality improvement	CNN	SRCNN	SR helps object detection in long-distance view.	PSNR:53.6; SSIM:0.977
	Almasri [65], 2018	Campus	ULB17-VT dataset; (512+58) pairs;	60×80	Super-Resolution		SRGAN	Visual-thermal fusion can enhance the thermal SR image quality	PSNR:52.4; SSIM:0.950
	Kwasniewska [31], 2020	Facial area	200 videos from 40 volunteers in 16 bits	41×41×1	Super-resolution	CNN	DRESNet	The performed evaluation showed that PSNR can be improved even by 60% if full 16-bit depth resolution data is used instead of 8 bits.	PSNR:53.3; SSIM:0.990
	Xu [33], 2020	Outdoor scenes	41 pairs of visible and infrared images[66]	128×128	Image Fusion	GAN, CNN	GAN, ResNet	GAN and ResNet improved the performance of visible and infrared image fusion.	SSIM 0.580
Medical	Guan [36], 2018	Breast	11 patient	640×480	Breast segmentation	CNN	C-DNN	Adding training samples from the same patient improves segmentation performance.	IoU:0.942
	Baffa [35], 2018	Breast cancer	Database for Mastology Research with Infrared Image - DMR	640×480	Classification	CNN	LeNet	Dynamic inspection data is used.	Accuracy: 0.980
	Kakileti [37], 2019	Breast	150 subjects at five different view angle	240x320, 640x 480	Segmentation	CNN	ResNet, V-Net	Breasts region detection independent of the image capture and view angle	Dice index:0.916

	Maldonado [39], 2020	Diabetic foot	DB1 contains 108 images from 17 volunteers with 4 different backgrounds and DB2 contained a total of 141 images from 47 new volunteers	240x320	Segmentation	CNN	mask R-CNN	DB2 is used for training and DB1 is used for validation.	Accuracy: 0.900
	Cruz-Vega [38], 2020	Diabetic foot	110 images	227×227×3	classification	CNN	DFTNet	Two traditional machine learning classifiers and three models based on CNNs were used for classification of the thermograms.	Accuracy: 0.945
Object detection & tracking	Ha [42], 2017	street scenes	1569 images	640×480	semantic segment	CNN	MFNet	Multi-spectral information is used.	Class average: 0.591; mean IoU: 0.649
	Mazur-Milecka [41], 2020	Laboratory animals	300 minutes of social behaviour recordings of rats	320 × 240	Semantic segmentation	CNN	U-Net, V-Net	Temperature range affects the segment results.	U-Net IoU: 0.915; V-Net IoU: 0.737
	Algarni [45], 2020	Outdoor and indoor scenes	1700 images	224×224	Template matching	CNN	VGG	CNN feature outperforms the HOG feature in object tracking.	Maximum achieved Accuracy: 1.00
	Zhang [67], 2020	Human and	RGBT210 [46] contains 210 aligned visible and	255×255×3,	Object tracking	CNN	Dynamic Siamese	Dynamic Siamese network is introduced for object tracking	Precision: 0.642

		cars in outdoor	infrared videos	127×127×3 for template			network	using the fusion feature of visible and infrared images.	
	Zhang [43], 2020	Outdoor and indoor scenes	Three public dataset [68-70] with over 846 image pairs	256×256×4	Saliency object detection	CNN	VGG-16	Multi-level CNN features have improved the performance on challenging conditions.	F-measure: 0.873
	Tu [44], 2020	Outdoor and indoor scenes	1000 aligned RGB-T images pairs [44]	480 × 640	Saliency object detection	Graphy learning	Collaborative graph learning	Collaborative graph learning is used for saliency detection.	F-measure: 0.724
	Huda [48], 2020	People in outdoor	PD-T[48] with 1950 annotated images	416×416×3	Object detection	CNN	YOLOv3	This paper highlight the importance of dataset variety in transfer learning.	F1-score: 0.902
	Ye [49], 2020	Night-time pedestrian	Three public datasets[71-73] with over 524318 visible and 13969 images images	288 × 144	Pedestrian retrieval	CNN	Two-stream CNN with ResNet50 as backbone	The integration of cross-modal feature distribution and contextual information with normal CNN has improved the performance in night-time pedestrian retrieval.	mAP: 0.463

Face recognition	Kwaśniewska [50], 2017	Face	68519 images	299×299	Face Detection & tracking	CNN	Inception v3	Short time (42ms) and high accuracy (89.2%).	Accuracy: 0.892
	Kopaczka [51], 2018	Face	2935 images of 90 subjects	1024 × 768	Face alignment &	CNN	deep alignment network (DAN)	A database with full manual annotations for 68 facial landmark points is proposed	Normalized point-to-point error: 0.04
Behavior detection	Kawashima [52], 2017	Human action	2,520 sequences	16×16	Action recognition	CNN, LSTM	CNN, LSTM	Deep Learning-based action recognition method that combines convolution layers and an LSTM layer for learning spatio-temporal representation	Accuracy: 0.911
	Imran [53], 2020	Sign Language	Four public visual dataset [74-77] with 6037 samples and one thermal dataset [53] with 1039 samples	224×224×3	Classification	Ensemble learning	MobileNet , KELM	Short videos are compressed into three images, and ensemble learning is used with pretrained MobileNet and KELM.	Accuracy: 0.772 (On IITR Sign Language Thermal Dataset [53])
Electrical equipment and machine	Gong [54], 2018	Electrical equipment	7955 images	480 × 480	Equipment detection	CNN	YOLO-based	predicts the coordinates, orientation angle, and class type of each equipment part	mAP: 0.932 (IoU=0.5)
	Ullah [55], 2020	High voltage	1075 defective and 925 non-defective images	224×224	Classification	CNN, Ramdom	AlexNet, Ramdom Forest	The combination of pretrained AlexNet and random forest outperforms LeNet and VGG.	F-measure: 0.950

		electrical equipment				Forest			
SHM	Ali [56], 2019	steel bridge	2000 images with a resolution of 299 × 299 pixels and 200 images with a resolution of 640 × 480	299 × 299	Subsurface damage detection	CNN	modified deep inception neural network (DINN)	The results are validated by ultrasonic pulse velocity tests	F1-score: 0.801
	Li [57], 2020	Rotating machine	6000 images	100×100×3	Fault classification	CNN	LeNet-based	The combination of CNN and thermography has better performance compared to the traditional vibration-based method.	Accuracy: 0.986
Remote sensing	Bigdeli [58], 2019	remote sensing	Multi-resolution	734 × 299 × 18	classification	CNN	CNN	CNN and SVM are combined for classification.	Accuracy: 0.951
Photovoltaic inspection	Dunderdale [59], 2020	Photovoltaic	398 singular defective and 400 singular non-defective PV modules	224×224×3	classification	CNN	VGG, MobileNet	The VGG-16 and MobileNet models are shown to provide good performance for the classification of defects.	Accuracy: 0.858 (VGG-16) Accuracy: 0.861 (MobileNet)
	Akram [60], 2020	Photovoltaic	Collected 893 images	100×100×1	classification	CNN	VGG-16 with transfer	The comparison shows that the develop-model transfer learning technique can help to improve	F1-score: 0.990

							learning	the performance.	
Quality Control	Izquierdo [61], 2020	Honey	78,129 images from 56 samples	228×228×3	classification	CNN	LeNet-like	Two-stage classifier, qualitative detection and quantitative detection of adulterant.	Accuracy: 0.950
	Izquierdo [62], 2020	Olive oil	5515 images from 24 thermographic videos	228×228×3	classification	CNN	LeNet-like	Two-stage classifier.	Accuracy: 0.970
	Zeng [63], 2020	Pears	4371 images from 300 pears	32×32×1	classification	CNN	LeNet	Basic classification method.	Accuracy: 0.990
Materials	Wu [64], 2020	Composites	1200 images from simulation	64×64×1	Semantic Segmentation	CNN	U-Net	Simulation data is used for training and testing.	Accuracy: 0.979

4. DL in active thermography NDT

Like any other non-destructive testing methods, the objective of active thermography is to detect and evaluate the defects or discontinuity in materials. Currently, the usage of DL in active thermography is fewer than that in passive thermography. As active thermography NDT always generates the thermal video, we divide DL in active thermography NDT into three categories based on the structure of data.

4.1 1D model

Due to the limitation of data, many scholars consider the thermogram sequence as the array of temperature series. Therefore, some network with the input of a vector is applied, such as RNN[78], 1D CNN[79], AE[80].

In [78], the thermal image sequence is regarded as the array of temperature series. Long short-term memory recurrent neural network (LSTM-RNN) model which automatically classifies common defects occurring including debonding, adhesive pooling, and liquid ingress in honeycomb materials. This LSTM-based algorithm has a greater than 90% sensitivity in classifying water, and hydraulic oil ingress. It has a greater than 70% sensitivity in classifying debonding and adhesive pooling.

In Ref. [79], 1D CNN model is applied to analysis the temperature series of CFRP specimen in LT setup. A novel two-stream CNN architecture shown in Fig. 13, is used to extract/compare features in a pair of 1D thermal signal sequences for accurate classification/differentiation of defective and non-defective regions. The experimental results reveal that the 1D CNN model has enhanced the contrast compared to principle component analysis (PCA) and Fourier transform.

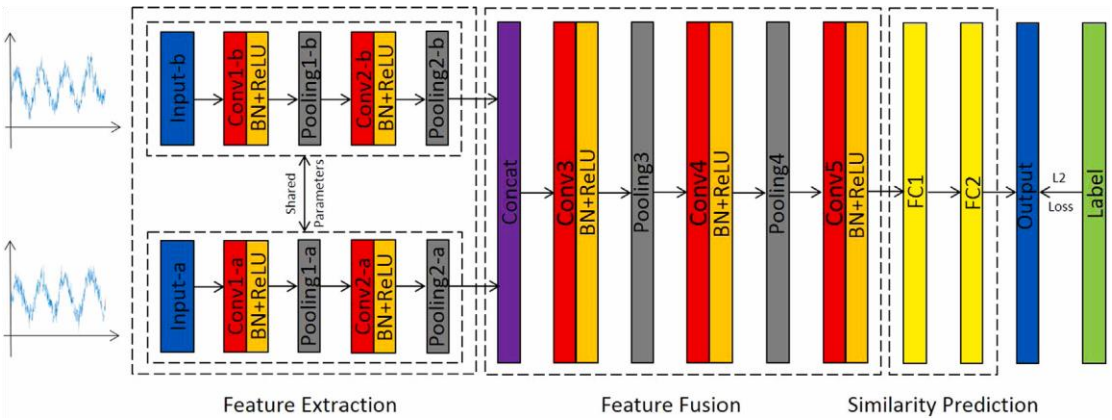

Fig. 13. Network architectures of Ref. [79] consisting of three major parts: feature extraction, feature fusion, and similarity prediction

In Ref. [80], automatic encoder thermography (AET) is proposed to improve the visibility of rear surface cracks during inductive thermography by employing the Autoencoder (AE) algorithm. The workflow of AET is illustrated in Fig. 14. Encoder is an important block to construct deep learning architectures. Through this framework,

underlying features of rear surface cracks are efficiently extracted and new clearer images are constructed.

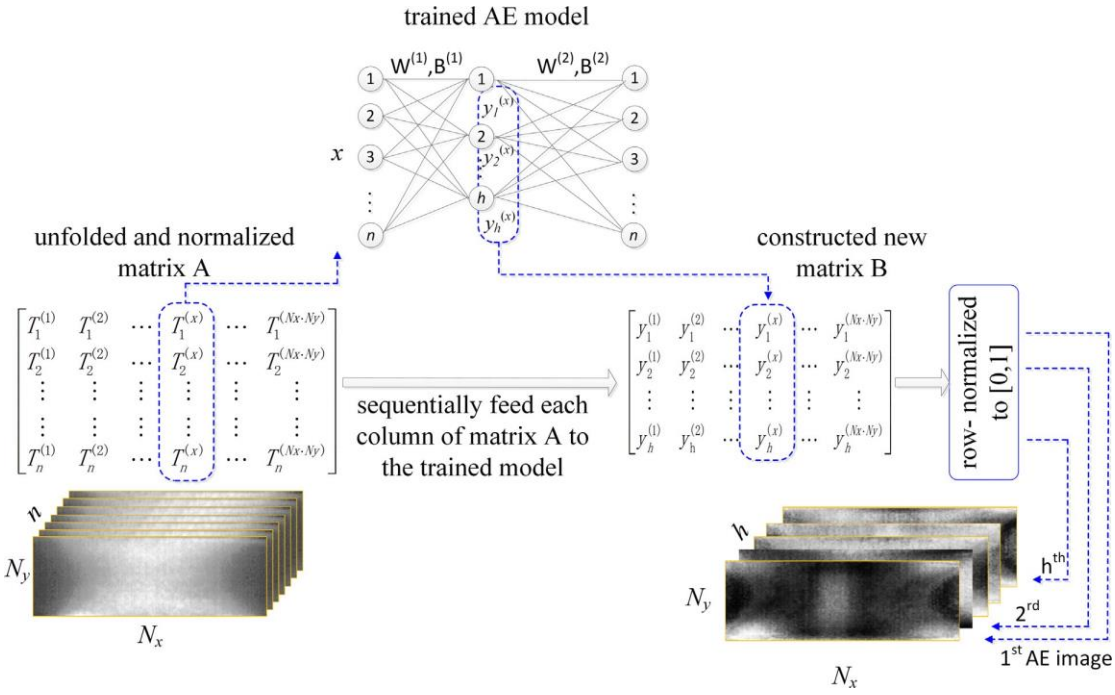

Fig. 14. Workflow of AET in Ref. [80]

4.2 2D model

2D models, such as CNN, has been widely used in VL images and passive IRMV for object classification, object detection, semantic segmentation, etc. However, the subsurface defects are usually revealed with the range of surface temperature, hindering the wide use of 2D model in active IRMV. 2D CNN is used for the inspection of surface crack [81, 82].

In Ref. [82], a deep learning–based autonomous concrete crack detection technique using hybrid images combining vision and infrared thermography images is proposed. Large-scale concrete-made infrastructures such as bridge and dam can be effectively inspected by spatially scanning the unmanned vehicle–mounted hybrid imaging system including a vision camera, an infrared camera, and a continuous-wave line laser. Therefore, the reconstructed images are divided into several patches and each patch is classify using pretrained GoogLeNet with transfer learning technique to achieve automated crack identification and visualization. The inspection result is visualized in Fig. 15. The proposed technique is experimentally validated using a lab-scale concrete specimen with cracks of various sizes. The test results reveal that macro- and microcracks are automatically visualized while minimizing false alarms.

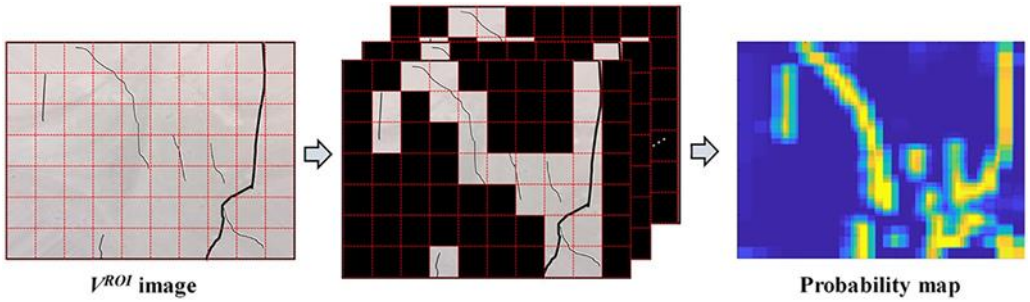

Fig. 15.The inspection result of Ref. [82]

4.3 3D model

3D model is the future direction of active IRMV, combining both temporal and spatial information. However, bas 3D data is difficult to process, there are many ways to simplify this operation, including deal temporal and spatial separately [83], compressing the temporal information [84], and integrated temporal-spatial network [85].

Based on the assumption that the temporal and spatial information of thermogram sequence are not related, a method combining YOLO and fully connected network (FCN) is proposed in Ref. [83]. The structure is demonstrated in Fig. 16. YOLO judges whether there are defects, and FCN predicts the depth of defects. The method can be seen as the integration of 1D model feed forward network and 2D model YOLO.

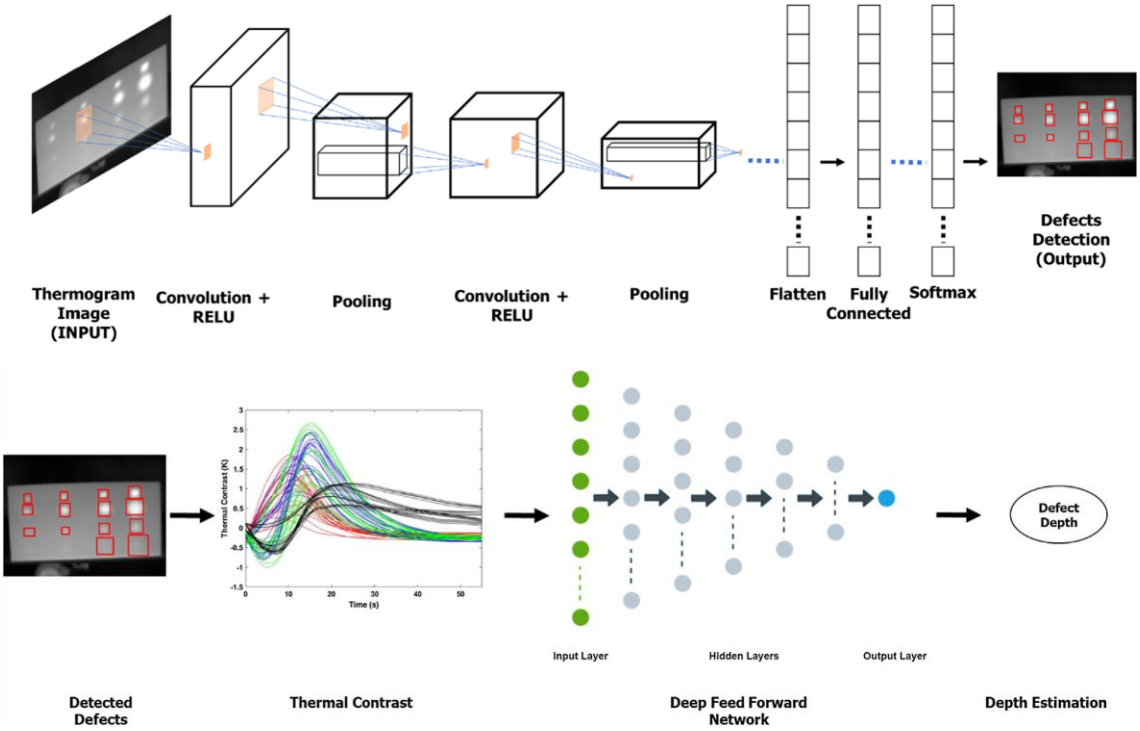

Fig. 16. Integration of YOLO and FCN in Ref. [83]

Because it is difficult to use deep neural network to process temporal and spatial information simultaneously, some scholars consider using non deep learning method to

compress temporal information, and then use deep learning method to process spatial information. In Ref. [84], an end-to-end pattern deep region learning structure shown in Fig. 17 is proposed to achieve precise crack detection and localization. The proposed structure integrates both time and spatial pattern mining for crack information with a deep region convolution neural network. PCA is used to extract the temporal information and compress the sequence into one image with more significant defect feature, while region proposal network Faster-RCNN is utilized to extract the defect in the image. Experiments on both artificial and natural cracks have shown attractive performance and verified the efficacy of the proposed structure.

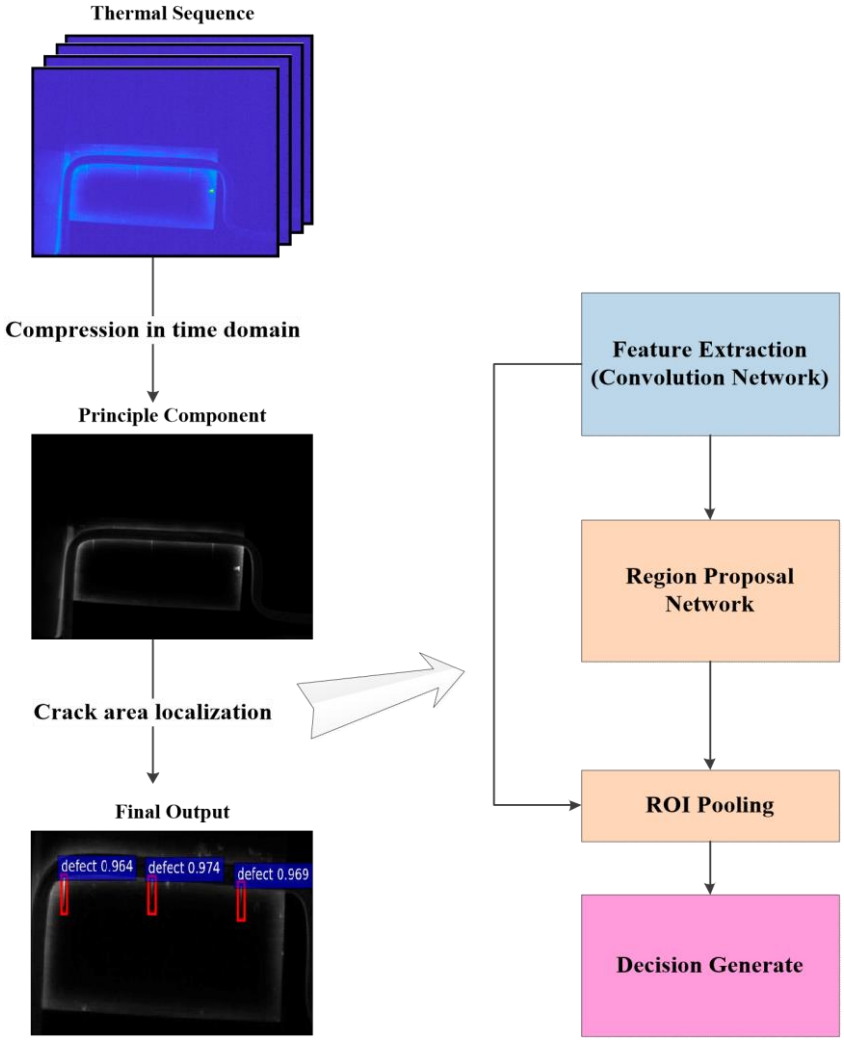

Fig. 17. Workflow of Ref. [84] consisting of temporal information compression and Faster R-CNN

Although it is very difficult, the processing of temporal and spatial information is integrated in Ref. [85]. A hybrid of spatial and temporal deep learning architecture for automatic thermography defects detection is proposed and shown in Fig. 18. Semantic segmentation network, visual geometry group-Unet (VGG-Unet) is applied to classify the defect region from the normal and a 3-layer long short term memory (3-layer LSTM) loop is applied to capture the physical properties that mining the feature of the temperature varies between the defect point and the non-defect point. The results show

that visual geometry group-Unet (VGG-Unet) cross learning structure can significantly improve the contrast between the defective and non-defective regions, and better evaluation results can be obtained after extracting certain features from low SNR images, especially of combining transient and spatial information.

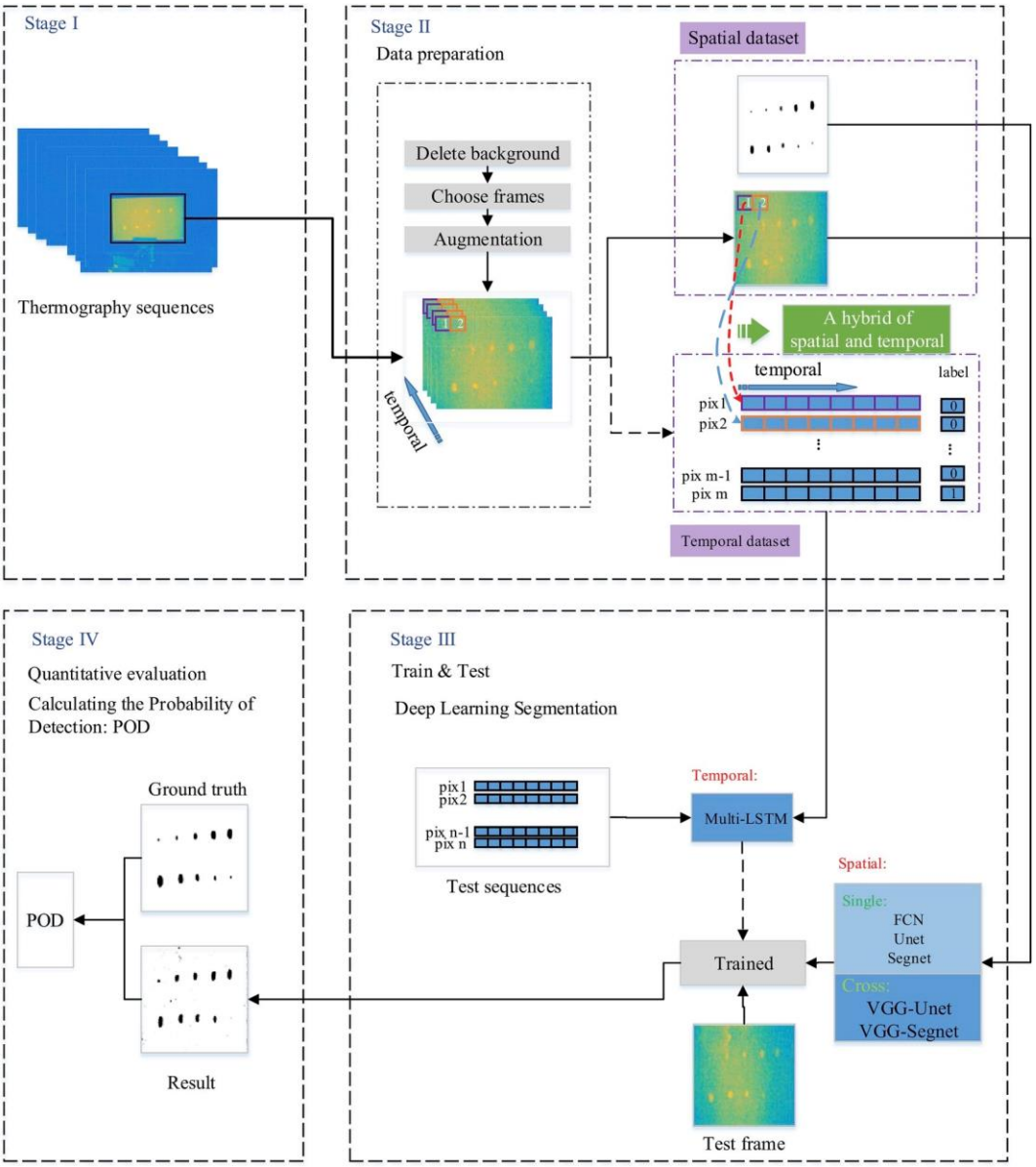

Fig. 18. Temporal and spatial strategy of Ref. [85] for segmentation of defect detectability

Table 3 DL in active thermography NDT

type	reference	objects	datasets	Stimulation	model
1D	Hu [78], 2019	Honeycomb-structured materials	320 × 240 series from 1 sample	Optical	LSTM-RNN
	Xie [80], 2018	Steel plate	3 samples	inductive	Automatic Encoder

	Cao [79], 2020	CFRP	380 × 380 series from 1 sample	LT	1D CNN
	Marani [86], 2018	CFRP	1 sample	Flash	Decision tree
	D'Orazio [87], 2005	Aircraft composites	1 sample	Lamp	MLP
	Marani [88], 2016	GFRP	1 sample	LT	Clustering
	Darabi [89], 2002	GFRP	1 sample	LT	MLP
2D	An [90], 2018	concrete crack	20K (augmented from 200) images	Laser	VGG-based CNN
	Yang [81], 2019	Metal crack	3K thermograms	Coil	Faster R-CNN
3D	Luo [85], 2019	composite	12 samples	Halogen lamp	VGG-Unet, LSTM
	Almasri [91], 2018	CFRP	8 samples	Optical	Wavelet-integrated alternating sparse dictionary matrix decomposition
	Hu [84], 2018	Metal	5 samples	ECPT	PCA, Faster-RCNN
	Saeed [83], 2019	CFRP	1 sample	Optical Pulse	YOLO, Fully Connected Network

5. Different Platforms for IRMV and DL

5.1 UAV-based IRT and DL

Frankly speaking, there are a lot of works on passive thermography with the help of UAV [92-94]. Here, we just want to talk the DL for the data collected by UAV carried thermal camera. In [95], the authors compare several methods for detecting failed solar panels, and found that drone detection is the most efficient. The thermographic images captured by drone are collected and various types of panel failures are explained. An examination by deep learning based on convolution neural network (CNN) revealed that single shot detection (SSD) is an effective method of automatic detection. The tests with the SSD model using 4-MW solar plant data are carried out and a mean average precision of 49.11% was achieved.

Some people has begun the UAV based active thermography. In [96], a micro-aerial vehicle (MAV) facilitated coating breakdown and corrosion assessment system

implementing deep learning for object recognition has been developed to provide effective coating breakdown and corrosion assessment for marine and offshore industries. In addition to the visual stereo camera, an infrared/thermal camera is set on the MAV. The visible and thermal images collected by the drone are transferred to integrated post-processing system, from where AI-based autonomous coating corrosion inspection results are sent to surveyor for result verification and report generation. In this work, TLCAF network is used to learn the region of interest for different types of coating failure. Bounding box is used for region of interest proposal, Faster RCNN framework and vgg19 model are used for CNN feature extraction, and SVM is used for feature classification. By using active thermography, it can identify corrosion behind coatings. This will greatly improve the work efficiency and reliability of coating inspection for surveyors.

5.2 Mobile-based IRT and DL

In recent years, some portable thermal cameras can be adapted to and mobile phones, which promote the development of TIRMV. In [97], an automatic detection of exercise-induced fatigue using thermal cameras adaptive to mobile phones has be proposed although the CNN used in their work is still running on a desktop computer. Fifty-seven hundred facial images has been collected via a Therm-App mobile thermal camera (19mm lens, 288x384 resolution, 8.7Hz, 17um wavelength), and then these images are trained and tested by using deep convolutional neural networks on a desktop computer featuring an Intel i7 processor, a Titan Xp GPU with 3840 CUDA cores running at 1.5GHz and Matlab 2015. The results have indicated that classification of fatigued cracks is possible. The CNN after training has obtained an accuracy over 80% during testing when utilizing single thermal images. In [98], an automatic method based on mobile phone carried FLIR One (160×120, 0.1 °C) and deep neural networks (DNNs) was proposed to detect anomaly for bus bar (meter), bus bar (circuit), IP phone and PC motherboard. The DNN model is trained to learn the statistical regularities of normal thermal images, and anomalies are detected based on pixel-wise comparison between the learned reference temperatures and the actual temperatures. Some people aim to build a low computational cost deep learning network for running on an embedded or mobile system. In [99], DeepBreath, a deep learning model which automatically recognises people's psychological stress level (mental overload) from their breathing patterns was proposed based on a low cost thermal camera (FLIR One). After thermal videos are captured using the camera attached to a smartphone placed at a maximum distance of 55 cm from a person's nostril area, the respiration tracking algorithm was proposed to recover one-dimensional breathing signals from the nostril ROI on the thermal videos. The architecture of proposed network based on CNN consists of two convolutional layers, two pooling layers and one fully connected layer. The model is trained and tested with data collected from people exposed to two types of cognitive tasks (Stroop Colour Word Test, Mental Computation test) with sessions of different difficulty levels. CNN reaches 84.59% accuracy in discriminating between two levels of stress and 56.52% in discriminating between three levels.

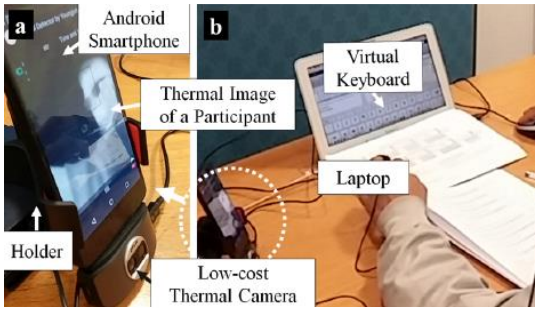

Fig. 19. Experimental setup based on a smart phone with a low-cost thermal camera [99].

5.3 Embedded-based IRT and DL

In [100], a complete integrated analysis system that can learn for the occurrence of water stress and can then recognize these preconditions in the real environment was proposed. The system uses a 300×128 resolution FLIR sensor connected via an interface to an integrated ARM Cortex A7 based system. The system allows the running of a deep learning neural network for both learning and testing. Following testing, the values (outputs from neural network) are transmitted using 866 MHz (LoRa) free radio communication. The system can make a local intelligent analysis and return different levels of alarm. Instead of transmitting the image to the central station on a conventional channel, the system conducts an on-site analysis over a period, identify potential water stress by recognizing templates, and will only convey indications of the crop state.

6. Trends

According to the above review, we can conclude the following trends:

1) The small amount of data used for training results that the trained models are specific to a particular scenario and lack generalizability. In addition to waiting for the price of infrared cameras to decrease, researchers should give priority to the fields that urgently need infrared cameras, such as target recognition in special scenes and defect detection.

2) In passive IRMV, DL of model fusion is an important topic in super-resolution, object detection and tracking, face recognition, and photovoltaics inspection. The fusion of VL and IRT shows its potential in security, identification, and quality inspection.

3) Cross modal identification that find corresponding relation from VL image library using the IR image will be widely used in some extreme occasions, such as non-invasive night-time monitoring and security.

4) In active IRMV, methods of analysing the temporal and spatial information of thermal data may vary, but the current technology is still not very efficient to cope with complex conditions. More generalized methods to reveal the spatio-temporal information for the inspection of defects should be proposed in the future.

5) Most of passive and active IRMV are at the stage of lab training and testing using

computers and there is still a gap between laboratory-based experiments and real industrial applications. With the growing compacity of edge computing, this problem can be solved in the future with embedded-based devices carried on robot, UAV or unmanned ship.

7. Conclusions

Thermal infrared machine vision is classified into passive thermography and active thermography based on the fact whether a controllable excitation source is required. Passive thermography is important supplement for conventional MV based on visible light camera and active thermography is an important non-destructive testing method for quality and safety. The development of deep learning and UAV makes IRMV more intelligent and automated. This paper introduces firstly and systematically the principle, camera, and data of IRMV, and then concludes and discusses the applications of DL in IRMV. After that, the development trends of IRMV and DL on various platforms like UAV, phone and embedded system are discussed. A mobile platform-based DL model is attracting in the future of TRIMV.

Acknowledgements

The work was supported by National Natural Science Foundation of China under Grant No. 61811530331 and Royal Society Newton Mobility Grant, IEC\NSFC\170387. The authors are also grateful to NSFC for sponsoring Dr Yunze He to the University of Surrey (UK) for joint research.

References

1. Buzmakov, A., et al., *Overview of Machine Vision Methods in X-ray Imaging and Microtomography*, in *Tenth International Conference on Machine Vision*, A. Verikas, et al., Editors. 2018.
2. Al-Mallahi, A., et al., *Detection of potato tubers using an ultraviolet imaging-based machine vision system.* Biosystems Engineering, 2010. **105**(2): p. 257-265.
3. Lemley, J., S. Bazrafkan, and P. Corcoran, *Deep Learning for Consumer Devices and Services Pushing the limits for machine learning, artificial intelligence, and computer vision.* Ieee Consumer Electronics Magazine, 2017. **6**(2): p. 48-56.
4. Akhtar, N. and A. Mian, *Threat of Adversarial Attacks on Deep Learning in Computer Vision: A Survey.* Ieee Access, 2018. **6**: p. 14410-14430.
5. Gopalakrishnan, K., et al., *Deep Convolutional Neural Networks with transfer learning for computer vision-based data-driven pavement distress detection.* Construction and Building Materials, 2017. **157**: p. 322-330.
6. Voulodimos, A., et al., *Deep Learning for Computer Vision: A Brief Review.* Computational Intelligence and Neuroscience, 2018.
7. Khodayar, F., S. Sojasi, and X. Maldague, *Infrared thermography and NDT: 2050 horizon.* Quantitative InfraRed Thermography Journal, 2016. **13**(2): p. 210-231.

8. Dyakonova, N., et al., *Terahertz vision using field effect transistors detectors arrays*. 2018 22nd International Microwave and Radar Conference. 2018. 711-14.
9. Yang, R. and Y. He, *Optically and Non-optically Excited Thermography for Composites: A review.* Infrared Physics & Technology, 2016. **75**: p. 26-50.
10. Nadjib Danial, A., *Using High Speed Shutter to Reduce Motion Blur in a Microbolometer*. 2013.
11. Rajic, N. and N. Street, *A performance comparison between cooled and uncooled infrared detectors for thermoelastic stress analysis.* Quantitative InfraRed Thermography Journal, 2014. **11**(2): p. 207-221.
12. Oswald-Tranta, B., M. Sorger, and P. O'Leary, *Motion deblurring of infrared images from a microbolometer camera.* Infrared physics & technology, 2010. **53**(4): p. 274-279.
13. Deng, B., et al., *Line Scanning Thermography Reconstruction Algorithm for Defects Inspection with Novel Velocity Estimation and Image Registration.* IEEE Sensors Journal, 2020.
14. Dionysopoulos, D., et al., *Imaging of barely visible impact damage on a composite panel using nonlinear wave modulation thermography.* NDT & E International, 2018. **95**: p. 9-16.
15. Dyrwal, A., M. Meo, and F. Ciampa, *Nonlinear air-coupled thermosonics for fatigue micro-damage detection and localisation.* Ndt & E International, 2018. **97**: p. 59-67.
16. He, Y., et al., *Volume or inside heating thermography using electromagnetic excitation for advanced composite materials.* International Journal of Thermal Sciences, 2017. **111**: p. 41-49.
17. Wang, Z., et al., *Comparative analysis of eddy current pulsed thermography and long pulse thermography for damage detection in metals and composites.* Ndt & E International, 2019. **107**.
18. Wang, Z., et al., *Image processing based quantitative damage evaluation in composites with long pulse thermography.* NDT & E International, 2018. **99**: p. 93-104.
19. Wang, D., et al., *Enhanced pre-processing of thermal data in long pulse thermography using the Levenberg-Marquardt algorithm.* Infrared Physics & Technology, 2019. **99**: p. 158-166.
20. He, Z., H. Wang, and Y. He, *Joint Scanning Laser Thermography Defect Detection Method for Carbon Fiber Reinforced Polymer.* IEEE sensors journal, 2020. **20**(1): p. 328-336.
21. West, J., D. Ventura, and S. Warnick, *Spring research presentation: A theoretical foundation for inductive transfer.* Brigham Young University, College of Physical and Mathematical Sciences, 2007. **1**(08).
22. Pan, S.J. and Q. Yang, *A survey on transfer learning.* IEEE Transactions on knowledge and data engineering, 2010. **22**(10): p. 1345-1359.
23. Li, W., et al. *Deep domain adaptive object detection: a survey*. in *2020 IEEE Symposium Series on Computational Intelligence (SSCI)*. 2020. IEEE.
24. Wang, M. and W. Deng, *Deep visual domain adaptation: A survey.* Neurocomputing, 2018. **312**: p. 135-153.
25. Li, W., et al., *Unsupervised Image-generation Enhanced Adaptation for Object Detection in Thermal images.* arXiv preprint arXiv:2002.06770, 2020.
26. Herrmann, C., M. Ruf, and J. Beyerer. *CNN-based thermal infrared person detection by*

domain adaptation. in *Autonomous Systems: Sensors, Vehicles, Security, and the Internet of Everything*. 2018. International Society for Optics and Photonics.

27. Iranmanesh, S.M., et al., *Deep Cross Polarimetric Thermal-to-visible Face Recognition.* arXiv: Computer Vision and Pattern Recognition, 2018.
28. Zhang, H., et al., *Synthesis of high-quality visible faces from polarimetric thermal faces using generative adversarial networks.* International Journal of Computer Vision, 2019. **127**(6-7): p. 845-862.
29. Chen, C. and A. Ross. *Matching Thermal to Visible Face Images Using a Semantic-Guided Generative Adversarial Network*. in *ieee international conference on automatic face gesture recognition*. 2019.
30. Zhang, H., et al., *Systematic infrared image quality improvement using deep learning based techniques*. SPIE Remote Sensing. Vol. 10008. 2016: SPIE.
31. Kwasniewska, A., et al., *Super-resolved thermal imagery for high-accuracy facial areas detection and analysis.* engineering applications of artificial intelligence, 2020. **87**.
32. Almasri, F. and O. Debeir. *Multimodal Sensor Fusion in Single Thermal Image Super-Resolution*. 2019. Cham: Springer International Publishing.
33. Xu, D., et al., *Infrared and Visible Image Fusion with a Generative Adversarial Network and a Residual Network*. applied sciences, 2020. **10**(2).
34. Hou, R., et al., *VIF-Net: An Unsupervised Framework for Infrared and Visible Image Fusion.* ieee transactions on computational imaging, 2020. **6**: p. 640-651.
35. Baffa, M.d.F.O. and L.G. Lattari. *Convolutional Neural Networks for Static and Dynamic Breast Infrared Imaging Classification*. in *2018 31st SIBGRAPI Conference on Graphics, Patterns and Images (SIBGRAPI)*. 2018. IEEE.
36. Guan, S., N. Kamona, and M. Loew. *Segmentation of Thermal Breast Images Using Convolutional and Deconvolutional Neural Networks*. in *2018 IEEE Applied Imagery Pattern Recognition Workshop (AIPR)*. 2018. IEEE.
37. Kakileti, S.T., G. Manjunath, and H.J. Madhu. *Cascaded CNN for View Independent Breast Segmentation in Thermal Images*. in *2019 41st Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*. 2019. IEEE.
38. Cruz-Vega, I., et al., *Deep Learning Classification for Diabetic Foot Thermograms.* sensors, 2020. **20**(6).
39. Maldonado, H., et al., *Automatic detection of risk zones in diabetic foot soles by processing thermographic images taken in an uncontrolled environment.* infrared physics & technology, 2020. **105**.
40. Porikli, F. and A. Yilmaz, *Object Detection and Tracking*. 2012.
41. Mazur-Milecka, M. and J. Ruminski, *Deep learning based thermal image segmentation for laboratory animals tracking.* quantitative infrared thermography, 2020: p. 1-18.
42. Ha, Q., et al. *MFNet: Towards real-time semantic segmentation for autonomous vehicles with multi-spectral scenes*. in *2017 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. 2017. IEEE.
43. Zhang, Q., et al., *RGB-T Salient Object Detection via Fusing Multi-Level CNN Features.* ieee transactions on image processing, 2020. **29**: p. 3321-3335.
44. Tu, Z., et al., *RGB-T Image Saliency Detection via Collaborative Graph Learning*. ieee transactions on multimedia, 2020. **22**(1): p. 160-173.

45. Algarni, A.D., *Efficient Object Detection and Classification of Heat Emitting Objects from Infrared Images Based on Deep Learning.* multimedia tools and applications, 2020. **79**: p. 1-24.
46. Li, C., et al., *Weighted Sparse Representation Regularized Graph Learning for RGB-T Object Tracking*, in *ACM Multimedia*. 2017. p. 1856-1864.
47. Tumas, P., A. Nowosielski, and A. Serackis, *Pedestrian Detection in Severe Weather Conditions*. ieee access, 2020. **8**: p. 62775-62784.
48. Huda, N.u., et al., *The Effect of a Diverse Dataset for Transfer Learning in Thermal Person Detection.* sensors, 2020. **20**(7).
49. Ye, M., et al., *Improving Night-Time Pedestrian Retrieval With Distribution Alignment and Contextual Distance.* ieee transactions on industrial informatics, 2020. **16**(1): p. 615-624.
50. Kwaśniewska, A., J. Rumiński, and P. Rad. *Deep features class activation map for thermal face detection and tracking*. in *2017 10th International Conference on Human System Interactions (HSI)*. 2017. IEEE.
51. Kopaczka, M., et al., *A Thermal Infrared Face Database With Facial Landmarks and Emotion Labels.* IEEE Transactions on Instrumentation and Measurement, 2018. **68**(5): p. 1389-1401.
52. Kawashima, T., et al. *Action recognition from extremely low-resolution thermal image sequence*. in *2017 14th IEEE International Conference on Advanced Video and Signal Based Surveillance (AVSS)*. 2017. IEEE.
53. Imran, J. and B. Raman, *Deep motion templates and extreme learning machine for sign language recognition.* the visual computer, 2020. **36**(6): p. 1233-1246.
54. Gong, X., et al., *A Deep Learning Approach for Oriented Electrical Equipment Detection in Thermal Images.* IEEE Access, 2018. **6**: p. 41590-41597.
55. Ullah, I., et al., *Deep Learning Image-Based Defect Detection in High Voltage Electrical Equipment.* energies, 2020. **13**(2).
56. Ali, R. and Y.-J. Cha, *Subsurface damage detection of a steel bridge using deep learning and uncooled micro-bolometer.* Construction and Building Materials, 2019. **226**: p. 376-387.
57. Li, Y., et al., *Rotating machinery fault diagnosis based on convolutional neural network and infrared thermal imaging.* chinese journal of aeronautics, 2020. **33**(2): p. 427-438.
58. Bigdeli, B., H. Amini Amirkolaee, and P. Pahlavani, *Deep feature learning versus shallow feature learning systems for joint use of airborne thermal hyperspectral and visible remote sensing data.* International Journal of Remote Sensing, 2019. **40**(18): p. 7048-7070.
59. Dunderdale, C., et al., *Photovoltaic defect classification through thermal infrared imaging using a machine learning approach.* progress in photovoltaics, 2020. **28**(3): p. 177-188.
60. Akram, M.W., et al., *Automatic detection of photovoltaic module defects in infrared images with isolated and develop-model transfer deep learning.* solar energy, 2020. **198**: p. 175-186.
61. Izquierdo, M., et al., *Convolutional decoding of thermographic images to locate and quantify honey adulterations.* talanta, 2020. **209**.
62. Izquierdo, M., et al., *Deep thermal imaging to compute the adulteration state of extra virgin olive oil.* computers and electronics in agriculture, 2020. **171**.
63. Zeng, X., et al., *Detection and classification of bruises of pears based on thermal images.*

postharvest biology and technology, 2020. **161**.
64. Wu, H., et al., *Deep learning-based reconstruction of the structure of heterogeneous composites from their temperature fields.* aip advances, 2020. **10**(4).
65. Almasri, F. and O. Debeir. *Multimodal Sensor Fusion In Single Thermal image Super-Resolution*. in *Asian Conference on Computer Vision*. 2018. Springer.
66. Alexander, T., *TNO Image Fusion Dataset*. 2014.
67. Choi, E.J., et al., *Development of Occupant Pose Classification Model Using Deep Neural Network for Personalized Thermal Conditioning.* energies, 2019. **13**(1).
68. Liu, T., et al., *Learning to Detect a Salient Object.* ieee transactions on pattern analysis and machine intelligence, 2011. **33**(2): p. 353-367.
69. Li, C., et al., *A Unified RGB-T Saliency Detection Benchmark: Dataset, Baselines, Analysis and A Novel Approach*. 2017.
70. Li, C., et al., *Weighted Low-Rank Decomposition for Robust Grayscale-Thermal Foreground Detection.* ieee transactions on circuits and systems for video technology, 2017. **27**(4): p. 725-738.
71. Zheng, L., et al., *MARS: A Video Benchmark for Large-Scale Person Re-Identification*, in *European Conference on Computer Vision*. 2016. p. 868-884.
72. Nguyen, D.T., et al., *Person Recognition System Based on a Combination of Body Images from Visible Light and Thermal Cameras.* sensors, 2017. **17**(3).
73. Wu, A., et al., *RGB-Infrared Cross-Modality Person Re-identification*, in *International Conference on Computer Vision*. 2017. p. 5390-5399.
74. Lin, Y.-C., et al., *Human action recognition and retrieval using sole depth information*, in *ACM Multimedia*. 2012. p. 1053-1056.
75. Liu, L. and L. Shao, *Learning discriminative representations from RGB-D video data*, in *International Joint Conference on Artificial Intelligence*. 2013. p. 1493-1500.
76. Donahue, J., et al., *Long-term recurrent convolutional networks for visual recognition and description*, in *Computer Vision and Pattern Recognition*. 2015. p. 2625-2634.
77. Ronchetti, F., et al., *LSA64: An Argentinian Sign Language Dataset*. 2016.
78. Hu, C., et al., *LSTM-RNN-based defect classification in honeycomb structures using infrared thermography.* Infrared Physics & Technology, 2019. **102**: p. 103032.
79. Cao, Y., et al., *Two-stream convolutional neural network for non-destructive subsurface defect detection via similarity comparison of lock-in thermography signals.* ndt & e international, 2020. **112**.
80. Xie, J., et al., *Improving visibility of rear surface cracks during inductive thermography of metal plates using Autoencoder.* Infrared Physics & Technology, 2018. **91**: p. 233-242.
81. Yang, J., et al., *Infrared Thermal Imaging-Based Crack Detection Using Deep Learning.* ieee access, 2019. **7**: p. 182060-182077.
82. Jang, K., N. Kim, and Y.-K. An, *Deep learning–based autonomous concrete crack evaluation through hybrid image scanning.* Structural Health Monitoring, 2019. **18**(5-6): p. 1722-1737.
83. Saeed, N., et al., *Automatic defects detection in CFRP thermograms, using convolutional neural networks and transfer learning.* Infrared Physics & Technology, 2019. **102**: p. 103048.
84. Hu, J., et al., *Pattern deep region learning for crack detection in thermography diagnosis*

system. Metals, 2018. **8**(8): p. 612.

85. Luo, Q., et al., *Temporal and spatial deep learning network for infrared thermal defect detection.* NDT & E International, 2019. **108**: p. 102164.
86. Marani, R., et al., *Modeling and classification of defects in CFRP laminates by thermal non-destructive testing.* Composites Part B: Engineering, 2018. **135**: p. 129-141.
87. D'Orazio, T., et al., *Defect detection in aircraft composites by using a neural approach in the analysis of thermographic images*. NDT & E International, 2005. **38**(8): p. 665-673.
88. Marani, R., et al. *Automatic detection of subsurface defects in composite materials using thermography and unsupervised machine learning*. in *2016 IEEE 8th International Conference on Intelligent Systems (IS)*. 2016. IEEE.
89. Darabi, A. and X. Maldague, *Neural network based defect detection and depth estimation in TNDE*. Ndt & E International, 2002. **35**(3): p. 165-175.
90. An, Y.-K., et al. *Deep learning-based concrete crack detection using hybrid images*. in *Sensors and Smart Structures Technologies for Civil, Mechanical, and Aerospace Systems 2018*. 2018. International Society for Optics and Photonics.
91. Almasri, F. and O. Debeir. *RGB Guided Thermal Super-Resolution Enhancement*. in *2018 4th International Conference on Cloud Computing Technologies and Applications (Cloudtech)*. 2018. IEEE.
92. Crusiol, L.G.T., et al., *UAV-based thermal imaging in the assessment of water status of soybean plants.* International Journal of Remote Sensing, 2019.
93. Ficapal, A. and I. Mutis, *Framework for the Detection, Diagnosis, and Evaluation of Thermal Bridges Using Infrared Thermography and Unmanned Aerial Vehicles.* Buildings, 2019. **9**(8).
94. Ortiz-Sanz, J., et al., *IR Thermography from UAVs to Monitor Thermal Anomalies in the Envelopes of Traditional Wine Cellars: Field Test.* Remote Sensing, 2019. **11**(12).
95. Y, H. and B. T. *Failure detection of solar panels using thermographic images captured by drone*. in *2018 7th International Conference on Renewable Energy Research and Applications (ICRERA)*. 2018.
96. Liu, L., et al. *An Integrated Coating Inspection System for Marine and Offshore Corrosion Management*. in *2018 15th International Conference on Control, Automation, Robotics and Vision (ICARCV)*. 2018.
97. Lopez, M.B., C.R. del-Blanco, and N. Garcia. *Detecting exercise-induced fatigue using thermal imaging and deep learning*. in *2017 Seventh International Conference on Image Processing Theory, Tools and Applications (IPTA)*. 2017.
98. Lile, C. and L. Yiqun. *Anomaly detection in thermal images using deep neural networks*. in *2017 IEEE International Conference on Image Processing (ICIP)*. 2017.
99. Cho, Y., N. Bianchi-Berthouze, and S.J. Julier *DeepBreath: Deep Learning of Breathing Patterns for Automatic Stress Recognition using Low-Cost Thermal Imaging in Unconstrained Settings*. arXiv e-prints, 2017.
100. Mazare, A.G., et al. *Embedded system for real time analysis of thermal images for prevention of water stress on plants*. in *2018 41st International Spring Seminar on Electronics Technology (ISSE)*. 2018.