Trends in Food Science & Technology 113 (2021) 193–204

Contents lists available at ScienceDirect

Trends in Food Science & Technology

journal homepage: www.elsevier.com/locate/tifs

Efficient extraction of deep image features using convolutional neural network (CNN) for applications in detecting and analysing complex food matrices

Yao Liu [a,c], Hongbin Pu [b,c,d], Da-Wen Sun [b,c,d,e,*]

[a] School of Mechanical and Electrical Engineering, Zhongkai University of Agriculture and Engineering, Guangzhou 510225, China
[b] School of Food Science and Engineering, South China University of Technology, Guangzhou 510641, China
[c] Academy of Contemporary Food Engineering, South China University of Technology, Guangzhou Higher Education Mega Center, Guangzhou 510006, China
[d] Engineering and Technological Research Centre of Guangdong Province on Intelligent Sensing and Process Control of Cold Chain Foods, & Guangdong Province Engineering Laboratory for Intelligent Cold Chain Logistics Equipment for Agricultural Products, Guangzhou Higher Education Mega Centre, Guangzhou 510006, China
[e] Food Refrigeration and Computerized Food Technology, University College Dublin, National University of Ireland, Agriculture and Food Science Centre, Belfield, Dublin 4, Ireland

ARTICLE INFO

Keywords:
Food detection
Convolutional neural network
Feature extraction
Deep learning
Food safety and quality

ABSTRACT

Background: The development of techniques and methods for rapidly and reliably detecting and analysing food quality and safety products is of significance for the food industry. Traditional machine learning algorithms based on handcrafted features normally have poor performance due to their limited representation capacity for complex food characteristics. Recently, the convolutional neural network (CNN) emerges as an effective and potential tool for feature extraction, which is considered the most popular architecture of deep learning and has been increasingly applied for the detection and analysis of complex food matrices.

Scope and approach: In the current review, the structure of CNN, the method of feature extraction based on 1-D, 2-D and 3-D CNN models, and multi-feature aggregation methods are introduced. Applications of CNN as a depth feature extractor for detecting and analyzing complex food matrices are discussed, including meat and aquatic products, cereals and cereal products, fruits and vegetables, and others. In addition, data sources, model architecture and overall performance of CNN with other existing methods are compared, and trends of future studies on applying CNN for food detection and analysis are also highlighted.

Key findings and conclusions: CNN combined with nondestructive detection techniques and computer vision system show great potential for effectively and efficiently detecting and analysing complex food matrices, and the features based on CNN show better performance and outperform the features handcrafted or those extracted by machine learning algorithms. Although there still remains some challenges in using CNN, it is expected that CNN models will be deployed on mobile devices for real-time detection and analysis of food matrices in future.

1. Introduction

With the improvement of the quality of life, people are increasingly conscious of high quality and safe food products in daily life, therefore the development of methods for reliably detecting and analysing food quality and safety is important for the industry (Liu et al., 2017).

Conventional analytical techniques and methods available for evaluating food quality are mainly destructive in nature, such as high-performance liquid chromatography, gas chromatography and gas chromatography-mass spectrometry. In addition, these techniques and methods are laborious, complex, and time-consuming, which requires highly skilled operators and a large number of chemical reagents (Cheng et al., 2017). Therefore, there is a great need to develop accurate, rapid, and advanced nondestructive methods for food evaluation and qualitative analysis (Elmarsy et al., 2012; Qu et al., 2015). With recent technological progress in photonics and optics, several spectroscopic techniques including near-infrared (NIR) spectroscopy, Raman spectroscopy and fluorescence spectroscopy have been attempted and

* Corresponding author. University College Dublin Belfield, Dublin, Ireland.
E-mail address: dawen.sun@ucd.ie (D.-W. Sun).
URL: http://www.ucd.ie/refrig, http://www.ucd.ie/sun (D.-W. Sun).

https://doi.org/10.1016/j.tifs.2021.04.042
Received 17 October 2020; Received in revised form 9 April 2021; Accepted 30 April 2021
Available online 6 May 2021

developed for the qualitative and quantitative analysis of food matrices (Ma et al., 2018; Wang et al., 2017).

In food detection and analysis, data acquisition is the first important step as the performance of food feature extraction highly depends on the status of the data (Teng et al., 2019). Many advanced nondestructive techniques are available for data acquisition including computer vision (Sun & Brosnan, 2003; Wang & Sun, 2003; Zheng et al., 2006), hyperspectral imaging (HSI) (Cheng et al., 2016b; Liu et al., 2018b; Ma et al., 2017; Pan et al., 2018), Terahertz imaging (Zhou et al., 2019) and so on. In addition, a number of food image datasets are also available such as Food-101, UECFood-100, UECFood-256, Food-524, Food-475, Food-5K and Food-11, which can be readily used for detecting and analysing food attributes (Ciocca et al., 2018; McAllister et al., 2018). Therefore, the main challenge is to design and develop accurate, efficient and objective algorithms to rapidly obtain food attributes.

There are many conventional machine learning algorithms used for food data analysis such as support vector machines (SVM) (Du & Sun, 2005), K-means clustering, and artificial neural networks (ANN) (Zhou et al., 2019). However, using these algorithms to exploit the food architecture information is a difficult task since foods are typically non-rigid and deformable objects, and foods generally exhibit high intra-class variance, i.e., multiple visual appearances within same food products and low inter-class variance, i.e., similar characteristics in different types of food products (Mezgec et al., 2017). Therefore, these traditional algorithms based on handcrafted features probably have poor performance due to their limited capacities in representing these complex characteristics (Situju et al., 2019).

Alternative algorithms that can overcome the above difficulties are those based on deep learning, also known as representational learning and feature learning (LeCun et al., 2015). For example, Gu et al. (2020) employed a dual-channel network to extract more complete image features and obtained good performance. Gu et al. (2021) studied the feasibility of ensemble meta-learning in better extracting the main features by fine-tuning the deep neural network. Deep learning is a derivative of artificial neural networks, aiming to establish the visual neural network of the human brain to analyze and process massive data, which can automatically extract data representations in a hierarchical way. Due to its feature learning and generalization ability, deep learning is widely applied for solving complex problems, which has achieved satisfactory performance in a variety of research areas such as medicine, forestry, and agriculture (Pouladzadeh et al., 2017).

Among the widely-used deep learning algorithms including convolutional neural network (CNN), fully convolutional network (FCN), stacked autoencoders (SAEs) and long short-term memory (LSTM), CNN and its derivative are the most popular model for processing visual-related problems including image classification, object detection and semantic segmentation (Teng et al., 2019). The operation of CNN, which is a type of highly parallelized method, is based on the principle of the forward and backward propagation algorithm that it can automatically learn to extract distributed features of input data in convolutional layers, and the deep features generated by CNN are more efficient and robust than features handcrafted (Ciocca et al., 2018). Therefore, recently CNN has been widely studied for extracting the features of foods. McAllister et al. (2018) employed the CNN model as deep feature extractors to capture the features of diverse food image datasets, which were subsequently fed into the machine learning algorithms to perform food classification. Teng et al. (2019) evaluated the performance of the feature based on CNN_5 architecture as compared with the Bag-of-Features (BoF) on a Chinese food image dataset, while Pan et al. (2020) utilized the combinational CNN with two-stream of subnets in parallel to extract different types of the features of food datasets, which were subsequently fusioned to improve the performance of classification by using a multi-feature aggregation method.

There are also a few relevant reviews published in the past. Kamilaris et al. (2018b) reviewed the applications of CNN in agriculture and indicate that CNN constitutes a promising technique with high performance for various agricultural problems, Zhou et al. (2019) focused on the applications of deep learning in the food domain, covering food identification, calories intake calculations, quality detection of fruits, vegetables, meat and aquatic products, food supply chain, and food contamination, indicated that deep learning as a promising tool outperforms manual feature extractors and machine learning algorithms, while most recently, Kumar et al. (2020) described several feature extraction methods based on conventional techniques and deep learning techniques for hyperspectral image classification, indicated that deep learning techniques were a better choice that suit the cubical form of the HSI data. However, despite the above reviews, no review is available on feature extraction methods related to different dimension based on CNN for food detection and analysis. Therefore, the objective of this review is to introduce feature extraction methods based on 1-D, 2-D, 3-D CNN models and discuss recent research advances in the food area including meat and aquatic products, cereals and cereal products, fruits and vegetables, and others. It is hoped that this review will provide critical information for further understanding of feature extraction techniques based on 1-D, 2-D, 3-D CNN and encourage more applications of CNN models for food detection and analysis.

2. Feature extraction using CNN

2.1. Fundamentals of CNN

The convolutional neural network represents a class of deep feed-forward neural networks, which is constructed by imitating the connective pattern of neurons in the human visual cortex (Hussain et al., 2019). A typical CNN architecture usually consists of convolutional (Conv) layers, activation layer, pooling (POOL) layer, and fully connected (FC) layer, as shown in Fig. 1. The Conv layers are made up of a series of convolution kernels, each of which obtains certain features from the input images, starting from basic features such as edges and shapes at initial layers and becoming more complex and specific features at the last layers. The parameters of convolution kernels (e.g. numbers, kernel size, strides, padding, etc.) should be adjusted and optimized according to the size of the input image and the architecture of the network (Zhou et al., 2019). Moreover, the neurons of the Conv layer are just sparsely connected to the neurons of adjacent layers, and individual neurons respond to the overlapping region of the receptive field by sharing weights until the entire visual area is covered (Hu et al., 2015; Mezgec et al., 2017). The activation layer behind Conv or FC layers is a non-linear operation of CNN models, which learns the non-linear representation of the output volume of the previous Conv or FC layer by a non-linear activation function such as Sigmoid, Tanh and rectified linear activation function (ReLU) (Paoletti et al., 2019). The POOL layer can reduce the spatial dimensions of the extracted feature maps and the number of network parameters by some non-linear numerical operations such as average-pooling, sum-pooling or max-pooling (Paoletti et al., 2019). For example, a max-pooling function divides the input data into non-overlapping rectangles and retains the maximum value for each region. After a set of Conv layers alternate with POOL layers, learning to abstract features of increasing levels, the FC layer as a classifier can be implemented at the end of modelling in order to classify input images into predefined labelled classes by integrating high-level features and information from previous layers (Kamilaris et al., 2018a). The output of FC layers represents the output of CNN models, and the number of output nodes depends on the number of labelled classes (Ciocca et al., 2018).

The early architecture of CNN is LeNet-5 (Lecun et al., 1998), which has been successfully employed for handwriting digital recognition. With the introduction of large scale labelled databases such as ImageNet, new deep learning technologies such as ReLU and Dropout, and new computing hardware such as graphics processing units (GPUs) and AlexNet (Krizhevsky et al., 2012) is the first notably framework that has gained excellent performance than traditional computer vision methods

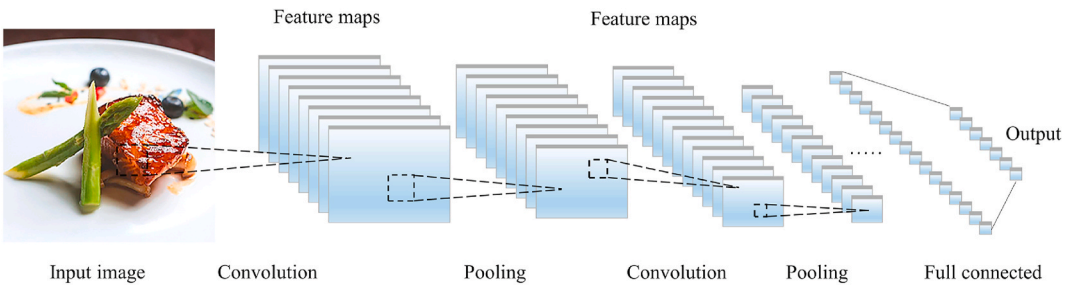


Fig. 1. A typical CNN architecture comprising convolutional (Conv) layers, pooling (POOL) layer, and fully connected (FC) layer for food detection and analysis.

for image recognition in ILSVRC (ImageNet Large Scale Visual Recognition Challenge). After the successful application of AlexNet, many deeper and wider network structures such as GoogleNet, VGGNet, Residual Networks (ResNet), and DenseNet have been proposed to handle complex problems of the restricted ability of feature representation (Pan et al., 2020).

With the rapid expansion of the depth and breadth of CNN models, more abstract and robust features can now be obtained by employing deeper networks. Jahani Heravi et al. (2018) designed a ConvNet architecture with 23 layers for food classification, which not only added the spatial pyramid pooling (SPP) layer to improve the performance of the model but also utilized bottleneck blocks to decrease the complexity of the model, realizing the parameters of the architecture fewer than ResNet and GoogleNet models. The parameters of the CNN architecture can be adjusted and updated in each training epoch of the forward and backward propagation process (Zhou et al., 2019). The forward propagation aims to calculate the error value between the output value and the labelled value according to the defined loss function, while the backward propagation intends to adjust the weight and bias of neurons based on this error value, thereby minimizing the loss function by using Stochastic Gradient Descent (SGD), Adaptive Gradient (AdaGrad), or Nesterov's Accelerated Gradient (NAG) algorithms (Mezgec et al., 2017).

All the above mentioned CNN architectures, including other derivative frameworks of CNN, have brought a series of eminent breakthroughs in the food area. However, one troublesome issue about training CNN models is that large-scale datasets are necessary as small-scale datasets (e.g. a few hundreds of images or less) might cause the overfitting problem (Pan et al., 2019). At present, four widely-used methods have been utilized to address the overfitting problem. The first skill is fine-tuning technology that takes trained weights and biases of the pre-trained model as the initialization value, and restart to train the model, as training the entire CNN model from scratch is complex and time-consuming (Pan et al., 2020). Ciocca et al. (2018) evaluated performances of the ResNet-50-S model trained from scratch and fine-tuned the ResNet-50 model on the Food-475 dataset, and found that the performance of fine-tuned ResNet-50 outperformed the ResNet-50-S. The second skill is the image augmentation technique for small and medium scale datasets. Pan et al. (2019) utilized four image augmentation techniques including rotation, flipping, translation and shearing to expand the size of the small-scale food dataset called MLC-41. The third skill is to use a CNN model pre-trained on a large-scale database as a depth feature extractor for small and medium scale datasets. The last skill is adjusting the network framework properly based on the basic CNN networks. Liu et al. (2018a) adjusted the AlexNet network architecture by adding a module called "Inception Module", which not only increased the depth of the network but also removed the computation bottlenecks. Therefore, the architecture and parameters of CNN should be correspondingly adjusted according to the specific food task. Fig. 2 summarizes the procedure of using CNN for food detection and analysis.

2.2. Extraction techniques

Feature extraction (FE) is usually considered as an indispensable step for image analysis and process, and it is a complicated and time-consuming process, which needs to determine the most appropriate types of features for the successful detection and analysis of certain food categories (McAllister et al., 2018). Generally, in the spectral feature extraction (FE) method, all pixel-wise spectra of the region of interest (ROI) in each sample are averaged as a spectral signature of ROI. At present, both features handcrafted using colour histogram, oriented gradients (HOG) histogram, scale-invariant feature transform (SIFT) and local binary pattern (LBP), and features captured using conventional methods including principal component analysis (PCA), wavelet transform (WT) and independent component correlation algorithm (ICA) can be utilized for food detection and analysis (McAllister et al., 2018; Zhou et al., 2019). However, using these features to perform food detection may have lower performance due to limited representation capacity for complex food characteristics (Situju et al., 2019).

On the other hand, efficient feature extraction methods based on CNN has been developed rapidly in recent years after the appearance of the ImageNet database, testifying the capacity of learning automatically high-level features. The features generated by CNN have been confirmed to be more efficient and robust than handcrafted and captured features (Ciocca et al., 2018; Jahani Heravi et al., 2018). Therefore, in order to extract deep features from pre-trained CNN, it is important to select a

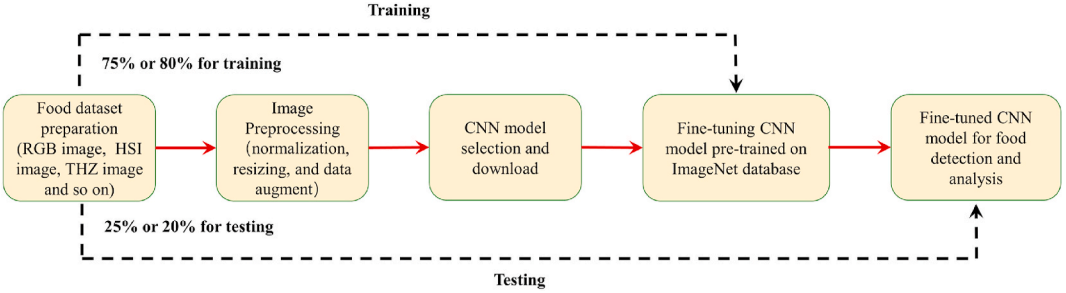


Fig. 2. A workflow for food detection and analysis. Five major steps are involved in this workflow including dataset preparation, image preprocessing, CNN model selection and download, fine-tuning CNN model, and application.

layer for each model. Generally, CNN extracts high-level features from the last output layer of the model (Pan et al., 2020). Additionally, it is worth mentioning that CNN models usually employ two-dimensional (2-D) convolution filters to analyze grayscale imagery and red-green-blue (RGB) images. However, one-dimensional (1-D) and three-dimensional (3-D) filters can be also employed to learn spectral features and spatial-spectral features, respectively (Ball et al., 2017; Zhou et al., 2019). In addition, the features extracted by 1-D, 2-D and 3-D CNN can be used to train another classifier like SVM and K-nearest neighbour (KNN). In addition, multi-feature aggregation techniques are also available for food detection and analysis. Fig. 3 compares the architectures of 1-D, 2-D and 3-D CNN. As shown in Fig. 3, 1-D, 2-D and 3-D CNN involve more complex model structure and computational efforts, however, it seems difficult for researchers with minimal background in computer science to program complex neural network models. Therefore, in order to reduce programming difficulty for researchers, many popular frameworks including Tensorflow, Keras, Theano, Pytorch and Caffe have emerged to help researchers quickly build 1-D, 2-D and 3-D CNN models by calling encapsulated function interfaces to stack some duplicate network layers such as Conv layer, POOL layer and FC layer (Zhou et al., 2019).

2.2.1. *One-dimensional spectral feature extraction*

Since the hierarchical architecture of CNN is designed to extract two-dimensional (2-D) image features, it is thus a challenge to employing CNN to extract one-dimensional (1-D) spectral signatures (Qiu et al., 2018). However, with the successful application of the CNN models, researchers start to develop 1-D CNN for signal processing (e.g., speech recognition and noise filtering), and the input of 1-D CNN can be regarded as a 1-D array (Al-Sarayreh et al., 2018). Therefore, 1-D CNN can be applied to capture spectral features of hyperspectral images with hundreds of contiguous spectral channels.

Fig. 3 (a) shows the 1-D CNN architecture using the feature extraction of a bruised apple as an example, which consists of an input layer, 1-D CNN block (stacking several Conv layers and POOL layers) and a fully-connected neural network (FNN) block. The average spectral pixels with a size of $n_{channels} \times 1$ are considered as input data (where$n_{channels}$ can be the number of original bands or the number of optimal bands selected), and there are m_1 1-D convolution kernels of size $k_1 \times 1$ in the first 1-D Conv layer, thereby the first 1-D Conv layer will generate m_1 feature maps of the size of $(n_{channels} - k_1 + 1) \times 1$ by 1-D convolution operation (Hu et al., 2015). Each feature map is obtained by taking the dot product between the weight matrix ω and the local area position x, and the value of a neuron V_{ij}^{x} at position x on the jth feature map in the ith layer can be evaluated by (Chen et al., 2016):

$$V_{ij}^{x} = \sigma\left(b_{ij} + \sum_{\mathrm{m}} \sum_{k=0}^{K_i - 1} w_{ij\mathrm{m}}^{k} V_{(i-1)\mathrm{m}}^{x+k}\right) \tag{1}$$

where $\sigma(\cdot)$ denotes the activation function of the ith layer, b_{ij} is an additive bias of jth feature map at the ith layer, m indexes the connection between the feature map in the $(i-1)$th layer and the current (jth) feature map, K_i is the width of the 1-D convolution kernel, and ω_{ijm}^{k} is a weight for input $V_{(i-1)\mathrm{m}}^{x+k}$ with an offset of k in 1-D convolution kernel.

The pooling process is triggered after the convolution stage, and it can make the features invariant from the location via decreasing the resolution of the feature maps. There are m_2 kernels of size $k_2 \times 1$ in the first POOL layer containing $m_2 \times n_3 \times 1$ nodes, and $n_3 = n_2/k_2$. Moreover, the neurons of the POOL layer connect a small $n \times 1$ patch of the Conv layer. The pooling operation of the max-pooling can then be completed using the equation below (Chen et al., 2016):

$$a_j = \max_{n \times 1}\left(a_i^{k \times 1} u(k, 1)\right) \tag{2}$$

where $u(k, 1)$ denotes a window function corresponding to the small $n \times 1$ patch of the Conv layer, anda_j is the maximum value of the

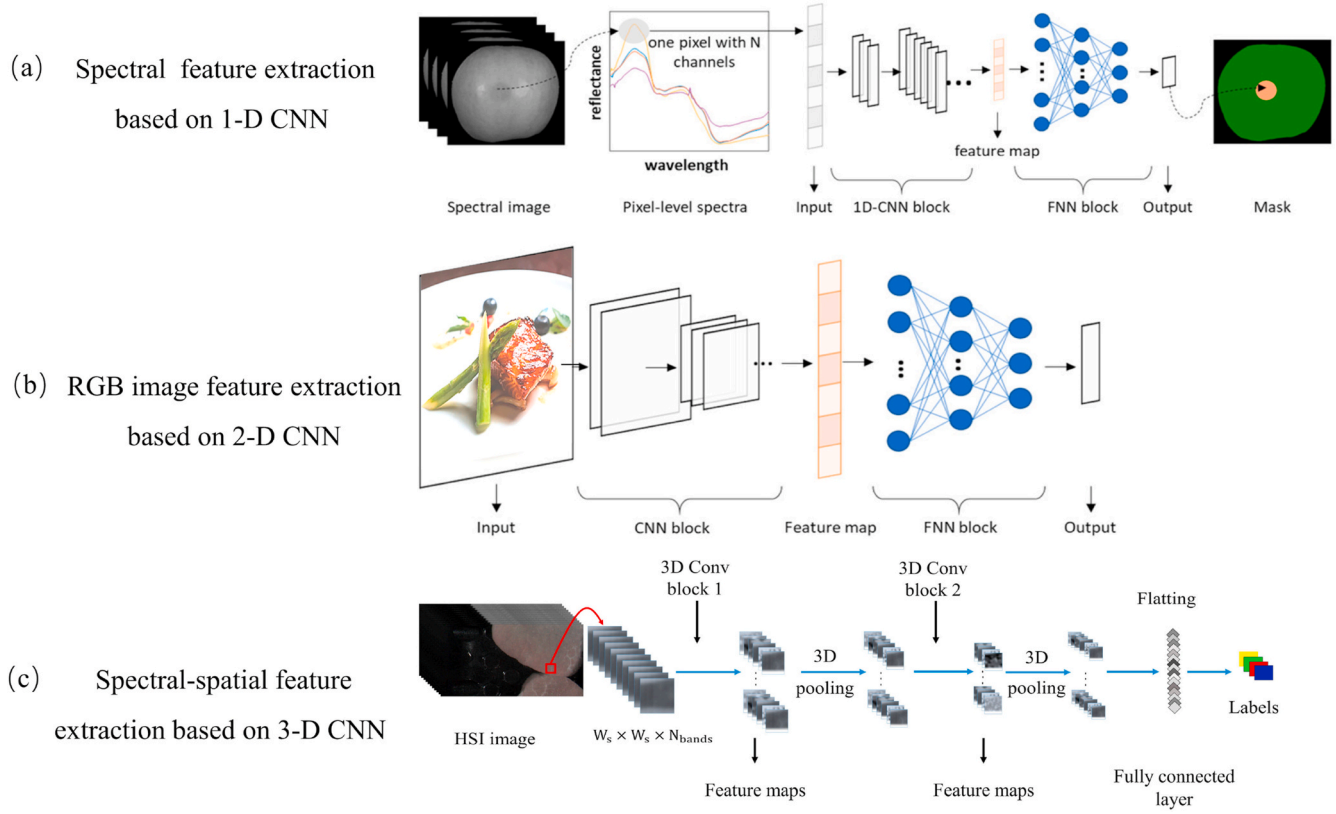


Fig. 3. The comparison for the feature extraction process of (a) 1-D CNN (Zhou et al., 2019), (b) 2-D CNN and (c) 3-D CNN (Al-Sarayreh et al., 2020).

neighbourhood.

After 1-D CNN block, the input pixel-level spectral vector can be transformed into feature vectors, which are employed to perform food detection and analysis, and obtain better performances than features handcrafted and features extracted by machine learning algorithm (Qiu et al., 2018; Wu et al., 2018). Moreover, the 1-D FE method can not only learn spectral features automatically but can also be used to explore potential feature combinations. In addition, the 1-D CNN model can be built quickly by using deep learning frameworks by calling encapsulated function interfaces including 1-D convolution filters, 1-D pooling, activation function, dropout algorithm and backward propagation algorithm. It is important to note that the input style of 1-D CNN is (N_{sample}, $N_{spectral\ bands}$, 1).

2.2.2. Two-dimensional image feature extraction

Two-dimensional CNN usually employs 2-D convolution filters to extract red-green-blue (RGB) food image features, and the architecture usually consists of several 2-D Conv layers and POOL layers. The Conv layer can extract features of the image by computing the response of the 2-D learning filter for the input image.

Fig. 3 (b) shows 2-D CNN architecture using the RGB feature extraction of a food image as an example. The size of the input image is just $N_1 \times N_1$, and there are m_1 2-D convolution kernels of the size $k_1 \times k_1$in the first 2-D Conv layer, thereby the first 2-D Conv layer will generate m_1 feature maps of the size $(N_1 - k_1 + 1) \times (N_1 - k_1 + 1)$ (depending on the size of the 2-D convolution kernels) by 2-D convolution operation. Moreover, each feature map is obtained by taking the dot product between the weight matrix ω and the local area position (x, y), and the value of a neuron V_{ij}^{xy} at position (x, y) on the jth feature map in the ith layer can be calculated by (Li et al., 2017):

$$V_{ij}^{xy} = \sigma\left(b_{ij} + \sum_{\mathrm{m}} \sum_{k=0}^{K_i-1} \sum_{p=0}^{P_i-1} w_{ijm}^{kp} V_{(i-1)\mathrm{m}}^{(x+k)(y+p)}\right) \tag{3}$$

where $\sigma(\cdot)$ denotes the activation function of the ith layer, b_{ij} is an additive bias of jth feature map at the ith layer, m indexes the connection between the feature map in the $(i-1)$th layer and the current (jth) feature map, K_i and P_i are the height and width of the 2-D convolution kernel respectively, and ω_{ijm}^{kp} is a weight for input $V_{(i-1)\mathrm{m}}^{(x+k)(y+p)}$with an offset of (k, p) in 2-D convolution kernel. The 2-D pooling process is implemented to decrease the resolution of the feature maps, and the implementation method of pooling is similar to 1-D CNN.

All layers of the 2-D CNN model can be stacked by calling encapsulated function interfaces of deep learning frameworks. However, it should be noted that the input style of 2-D CNN is (N_{sample}, W_s, W_s, $N_{channels}$). The 2-D CNN feature extraction method has been widely studied for extracting red-green-blue (RGB) food images. Mezgec et al. (2017) and Pandey et al. (2017) employed 2-D CNN architecture named NutriNet and ensemble architecture as a feature extractor to capture features of RGB food images, respectively, and their results showed that the 2-D CNN architectures were able to extract food image features of intra-class and inter-class variance. In addition to feature extraction of RGB food images, 2-D CNN can also be adapted to extract spatial features of hyperspectral images. Steinbrener et al. (2019) presented a 2-D GoogLeNet architecture, which was fine-tuned by adding multiple additional convolutional layers to make hyperspectral images similar to RGB images (projecting the hyperspectral data with 16 bands into three-band images) and performed feature extraction using a 2-D convolution kernel. Additionally, with respect to hyperspectral images, PCA, Pseudo-RGB model, kernel model and other methods can also be employed to compress information of a hyperspectral image, making hyperspectral data similar to RGB images.

2.2.3. Three-dimensional spatial-spectral feature extraction

Although 1-D and 2-D CNN feature extraction (FE) methods can effectively capture the spectral features and spatial features, it is impossible to extract joint spatial-spectral correlation features adequately using individual 1-D and 2-D CNN. Therefore, 3-D CNN can utilize 3-D convolution kernels to simultaneously capture the spatial-spectral features without any preprocessing or post-processing (Al-Sarayreh et al., 2020).

Fig. 3 (c) shows 3-D CNN architecture using the feature extraction of red meat as an example. A hyperspectral image cube with a size $W_S \times W_S \times M$ is considered as input data, where $W_S \times W_S$ and M represent spatial size (window size) and the number of original spectral bands, respectively, and there are m_1 3-D convolution kernels of the size $k_1 \times k_2 \times k_3$ in the first 3-D Conv layer, thereby the first 3-D Conv layer will generate m_1 feature maps of the size $(W_S - k_1 + 1) \times (W_S - k_2 + 1) \times (M - k_3 + 1)$ by 3-D convolution operation. Moreover, each feature map is obtained by taking the dot product between the weight matrix ω and the local area position (x, y, z), and the value of a neuron V_{ij}^{xyz} at position (x, y, z) of the jth feature cube in the ith layer can be determined by (Qi et al., 2019):

$$V_{ij}^{xyz} = \sigma\left(b_{ij} + \sum_{\mathrm{m}} \sum_{k=0}^{K_i-1} \sum_{p=0}^{P_i-1} \sum_{\mathrm{q}=0}^{Q_i-1} w_{ijm}^{kpq} V_{(i-1)\mathrm{m}}^{(x+k)(y+p)(z+q)}\right) \tag{4}$$

where σ denotes the activation function of the ith layer, b_{ij} is an additive bias of jth feature map in the ith layer, m indexes the connection between the feature map in the previous layer ($(i-1)$th layer) and the current feature map, K_i and P_i are the height and width of the convolution kernel in the spatial dimension, respectively, Q_i is the kernel size of spectral dimension, and ω_{ijm}^{kpq} is a weight for input $V_{(i-1)\mathrm{m}}^{(x+k)(y+p)(z+q)}$ with an offset of (k, p, q)in 3-D convolution kernels.

The construction of the 3-D CNN model is similar to 1-D and 2-D CNN. However, it is important to note that the input style of 3-D CNN is (N_{sample}, W_s, W_s, $N_{spectral\ bands}$, 1). The 3-D feature extraction method has been proved to be very effective to simultaneously capture the 3-D feature cubes of the spatial and spectral dimensions by applying 3-D kernels to hyperspectral data (Al-Sarayreh et al., 2020). Compared with the 1-D and 2-D convolution operation for a hyperspectral image, the 3-D convolution operation can alleviate spectral distortion and extract more complex three-dimensional features such as spatial-spectral correlation signatures and absorption differences between bands (Audebert et al., 2019). Moreover, the 3-D CNN model is well-suited theoretically to extract a 3-D feature cube for hyperspectral image classification since hyperspectral data are usually presented as a 3-D cube.

2.2.4. Multi-features aggregation

Since foods are typically non-rigid and deformable objects, which makes the process of exploiting their architectural information even more difficult. Moreover, foods generally tend to exhibit high intra-class and low inter-class variance (Mezgec et al., 2017). For food images with larger geometric variants, the features extracted and represented from the whole image by using Conv layer of single-stream CNN architecture often fail to achieve optimal recognition and classification performances (Jiang et al., 2020). Therefore, multi-features aggregation methods have been developed to improve the performances, and these aggregation methods use batch normalization (BN), auxiliary classifier and simple concatenation to aggregate different types of features from two or more CNN subnetworks due to the aggregation features being more complementary and discriminative (Pan et al., 2020; Pandey et al., 2017).

Batch normalization (BN) can be applied to provides zero mean and unit variance input to any layer of the neural network, and can be utilized to any set of activations in the network. Pan et al. (2020) presented the following equation for the linear transformation of mini-batch B with the size of m:

$$Y \leftarrow \gamma \frac{X - \mu_B}{\sqrt{\sigma_B^2 + \varepsilon}} + \beta \tag{5}$$

where$X = \{x_{1\ldots m}\}$ is the input of any layer, $Y = \{y_{1\ldots m}\}$ represents the output of linear transformations, μ_B and σ_B^2 represent mean and variance of the mini-batch, respectively.

For each activation x_i, there are pairwise parameters γ and β, which scale and shift the normalized value, and ε represents a constant, which is added to the mini-batch variance for numerical consistency. Additionally, with respect to aggregating the features from multiple subnetworks, the BN layer is usually placed before the feature fusion of the multiple subnetworks for avoiding the information loss of any subnetwork and the overfitting. This method also eliminates the distribution differences of the subnetworks in the procedure of aggregation and accelerates convergence (Pan et al., 2020). After batch normalization, an auxiliary classifier such as SVM, random forest, and fully-connected neural network (FNN) can then be utilized to concatenate the subnetworks for classification.

Several studies have been conducted to fuse multiple features for food classification. Jiang et al. (2020) proposed a multi-scale multi-view features aggregation (MSMVFA) method based on two 2-D CNN models, as shown in Fig. 4 (a). The ingredient information and category of food as the supervised label were employed to extract mid-level attribute features, high-level semantic features and deep visual features for each scale by fine-tuning the 2-D multi-scale ingredient and category CNN, respectively. Then three different types of features were fused into a multi-scale representation using the fusion method of normalization and simple concatenation. Finally, multi-view feature aggregation methods of z-score normalization and simple concatenation were further used to aggregate multi-scale features of three different types into the final representation. In contrast, Al-Sarayreh et al. (2018) proposed ensemble architecture of 1-D and 3-D CNN models, as shown in Fig. 4 (b). The 1-D and 3-D CNN models were utilized to extract the spectral feature and spatial features, respectively, and these features were then aggregated into the final joint spatial-spectral feature by using a fully connected layer.

The above studies showed that the multi-features aggregation method can rapidly and effectively coordinate the training of subnetworks and extract richer and more robust features. Moreover, features aggregation of two-stream or more subnets is more complementary and discriminative to geometrical deformation with respect to the single-stream CNN, which can also significantly improve the generalization capacity of the CNN model (Pan et al., 2020).

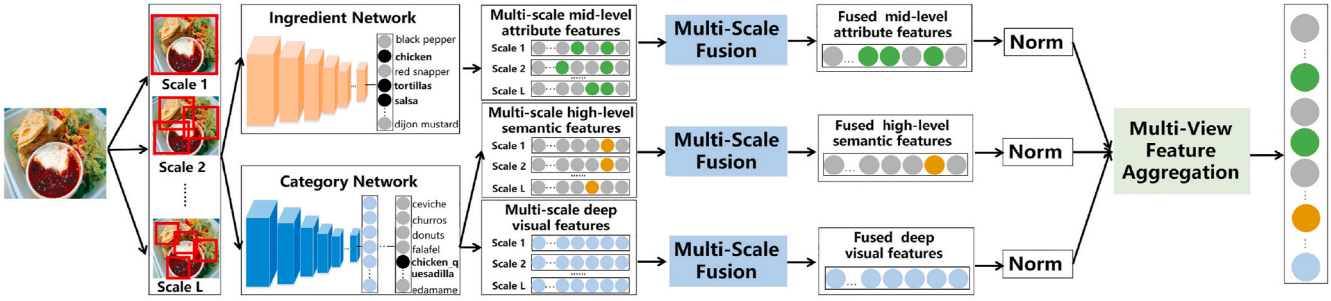
3. Applications for food detection and analysis

CNN models have been widely introduced into the food field for food detection and analysis, with the dominant framework being AlexNet, VGGNet, GoogleNet, Residual Networks (ResNet) and DenseNet. Table 1 summarizes the applications of 1-D, 2-D and 3-D CNN models for the detection and analysis of food products including meat and aquatic products, cereals and cereal products, and fruits and vegetables.

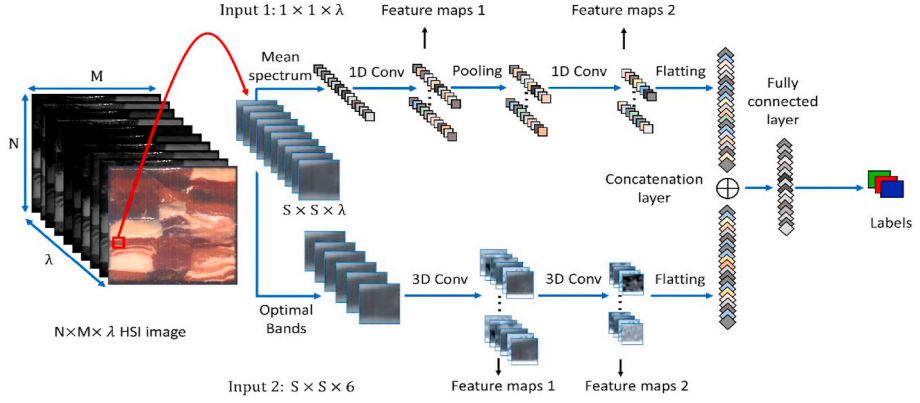
3.1. Meat and aquatic products

Meat and aquatic products are the first-choice source of animal protein supply for human diets (Cheng et al., 2016a). However, adulteration occurs for unlawful profits by adding low-quality products to the high ones. Therefore, rapid and reliable quality detection is of the utmost importance for consumers and distributors (Cheng et al., 2018; Jackman et al., 2011). CNN as a novel, effective, and powerful feature extraction tool can not only extract spectral features but also image features, and have been applied to the detection of meat and aquatic products (Al-Sarayreh et al., 2018; Xu et al., 2018).

Al-Sarayreh et al. (2018) employed 1-D and 3-D CNN subnets to extract the mean spectral feature of ROI and spatial features of six



(a) Multi feature aggregation based on same type of CNN, 2-D CNN and 2-D CNN



(b) Multi feature aggregation based on different type of CNN, 1-D CNN and 3-D CNN

Fig. 4. The comparison for multi-feature aggregation based on (a) two 2-D CNN subnetworks (Jiang et al., 2020) and (b) 1-D and 3-D CNN subnetworks (Al-Sarayreh et al., 2018).

Table 1
Applications of 1-D, 2-D and 3-D CNN models in food quality detection.

Products	Samples	CNN models	Results	References
Meat and fishery products	Fresh unpacked, fresh packed, frozen unpacked, frozen packed and frozen-thawed unpacked red meats	1-D and 3-D CNN	Classification rate of 94.4%	Al-Sarayreh et al. (2018)
	Fresh lamb, beef and pork	3-D CNN	Classification rates of 96.9% and 97.1% for NIR and VIS snapshot HIS, respectively	Al-Sarayreh et al. (2020)
	Wholesome and defective salmon fillets	2-D AlexNet	Classification rates of 92.7% and 91.6% for cross-validation and test sets, respectively	Xu et al. (2018)
	Most fresh, fresh, fairly fresh and spoiled carps	2-D VGG-16	Classification rate of 98.21%	Taheri-Garavand et al. (2020)
Cereals and cereals products	Japonica and Indica rices	1-D VGGNet	Classification rates of 89.6% and 87.0% for training and testing sets, respectively	Qiu et al. (2018)
	Medium-grain, round-grain and long-grain rices	2-D DCNN	Classification rates of 99.4% and 95.5% for calibration and validation sets, respectively	Lin et al. (2018)
	Qualified and defective maize kernels	2-D ResNet	Prediction accuracy of 98.2%	Ni et al. (2019)
	Soybeans in different varieties	1-D CNN	Classification rate of each variety over 90%	Zhu et al. (2019)
	Normal wheat and those with impurities (insects, stalks, grass, awns, spikelet)	2-D WheNet	Recognition rates of 98.59% (Top-1%) and 99.98% (Top-5%) for testing set	Shen et al. (2019)
	Barley in different varieties	2-D CNN	Classification rate of 93.21%	Kozlowski et al. (2019)
	Bread crusts with different baking periods (0, 5, 10, 15, 20, 25 and 30 min)	2-D Short-CNN	Recognition rate of 98.8%	Cotrim et al. (2020)
Fruits and vegetables	Olives in different varieties	2-D Inception-ResnetV2	Classification rate of 95.91%	Ponce et al. (2019)
	Dates in different varieties and maturity stages	2-D VGG-16	Classification rates of 99.01%, 97.25%, and 98.59% for varieties, maturity, and harvesting decision, respectively	Altaheri et al. (2019)
	Normal and defective apples	2-D CNN	Classification rate of 96.5%	Fan et al. (2020)
	Mealy and non-mealy apples	2-D AlexNet	Classification rate of 91.11%	Lashgari et al. (2020)
	Sound and damaged blueberries	3-D ResNet	Average accuracy of 88.44%	Wang et al. (2018)
	Sound and decayed blueberries	3D-CNN	Detection rate of 92.15%	Qiao et al. (2020)
	Healthy and bruised jujubes from different geographical origins	1-D CNN	Detection rate of over 85% for Vis-NIR spectra from all geographical origins	Feng et al. (2019)
	Healthy and damaged sour lemons	2-D CNN	Classification rate of 100%	Jahanbakhshi et al. (2020)
	Plums in different varieties	2-D AlexNet	Classification rates from 91% to 97%	Rodriguez et al. (2018) da
	Healthy and defective tomatoes	2-D ResNet	Classification rate of 96.2%, 94.2% and 94.6% for training set, validation set and testing set, respectively	Costa et al. (2020)
Others	Normal milk and milk samples adulterated with sucrose, soluble starch, sodium bicarbonate, hydrogen peroxide, and formaldehyde	1-D CNN	Classification rates of 98.76% and 96.95% for binary and multiclass, respectively	Asseiss Neto et al. (2019)
	Normal sesame oil and sesame oil counterfeited with rapeseed oil, soybean oil and corn oil	2-D AlexNet	$R^2 > 0.99$, RMSEP of 0.99%, 2.20% and 1.64% for three counterfeit samples, respectively	Wu et al. (2020)

optimal bands and aggregated these features to form the final joint spatial-spectral features based on the feature fusion method of fully-connected neural network (FNN), which was then used for adulteration detection of five different states of mixed red meats. They showed that the joint features performed better than the spectral and texture feature based on SVM and obtained an overall detection accuracy of 94.4%. Recently, Al-Sarayreh et al. (2020) further proposed a 3-D CNN model to simultaneously capture the spectral and spatial feature of sample for the classification of three different meat species using snapshot HSI and line-scanning HSI system and found that the 3-D CNN model was excellent on the snapshot HSI system in near-infrared (NIR) range and visible (VIS) range with a classification accuracy of 96.9% and 97.1%, respectively.

With regard to aquatic products, surface defects are considered the main issue in downgrading the quality of the products. Xu et al. (2018) utilized a 2-D AlexNet model and histograms of oriented gradients (HOG) model to extract the features of salmon fillet samples, which were subsequently fed into SVM for defect detection of the fillets. A classification accuracy of 80.0% was obtained based on handcrafted features of HOG, however, the use of the 2-D AlexNet feature was obviously better than the handcrafted features of HOG, resulting in a classification accuracy of 91.6% for testing sets. In addition, freshness detection of aquatic products is also of great concern (Dai et al., 2016). Taheri-Garavand et al. (2020) employed a 2-D VGG-16 architecture as a deep feature extractor to extract the RGB image features of common carp with four different levels of freshness, which were subsequently fed into a classifier block for freshness detection. Their results indicated that the performance of the VGG-16-based feature was superior to traditional feature extraction methods, and a classification accuracy of 98.21% was achieved.

The above studies confirm that CNN features are more efficient and robust as compared with features handcrafted and features captured based on traditional algorithms, therefore CNN features provide the most useful information of samples and can achieve encouraging results. In addition, CNN coupled with other nondestructive detection techniques should be further explored to detect the quality of more meat and aquatic products in different conditions such as during transportation, for various storage periods or at varying temperatures, which can be a focus for future studies.

3.2. Cereals and cereal products

Cereals including rice, maize, soybean, wheat and barley are important agricultural products but they are vulnerable to resist diseases and insects. Therefore, the development of effective detection methods for guaranteeing the quality and yield of cereals are of great interest. CNN combined with computer vision systems or hyperspectral imaging techniques can provide an efficient tool for detecting and analyzing cereals and cereal products (Kozlowski et al., 2019; Zhu et al., 2019).

Rice from different growth environments has different nutrients and flavours (Li et al., 2020; Qiu et al., 2018), knowing the species of rice is thus important for farmers and consumers. Qiu et al. (2018) utilized a modified 1-D VGGNet model with different sizes of training sets to directly extract the 1-D spectral feature of four varieties of rice from

hyperspectral images in two spectral ranges of 441–948 nm and 975–1646 nm and found that the classification accuracy gradually increased with increasing of the sample number of the training set and VGGNet model with 3000 training samples outperformed KNN and SVM models for 1-D spectra, which yielded the best classification accuracy of 87.0% in the spectral range of 975–1646 nm. Lin et al. (2018) combined a 2-D CNN architecture with a machine vision system to differentiate three varieties of rice and the image features of rice samples were extracted using CNN. In comparison with handcrafted features, the CNN model achieved the highest classification accuracy of 99.4% and 95.5% in the calibration and validation sets, respectively. The above studies demonstrated the possibility that 1-D and 2-D CNN combining hyperspectral imaging and machine vision technique can discriminate against species of rice with acceptable precision and accuracy.

On the other hand, Zhu et al. (2019) explored the feasibility of 1-D CNN models with a few training samples for identifying three varieties of soybeans with the aid of pixel-wise spectra of hyperspectral imaging (HSI) and found that the pixel-wise CNN model yielded the most satisfactory result. Furthermore, they concluded that the performance could be further improved by increasing the number of training set samples.

Maize is known as the "queen of the cereals" with high nutritional values and harvested as human food and livestock feed. However, the defect of maize kernels can degrade the value of maize, thereby defect detection of maize is needed. Ni et al. (2019) developed an automatic inspection machine based on a dual-side camera and 2-D ResNet model for defect detection of maize kernels. A total of 2040 pairs of maize kernel images were prepared and pretreated using dual-side camera and k-means clustering guided-curvature method, which were subsequently sent to 2-D ResNet for defect detection, achieving the best classification accuracy of 98.2%.

Besides studies focusing on quality detection of rice, soybean and maize, some attempts have also been made to quality detection of wheat and barley. In one study, Shen et al. (2019) proposed a 2-D WheNet convolutional neural network to extract the image features of normal wheat and five kinds of wheat with impurities for the detection of the impurity in wheat samples. Compared with the detection results of ResNet_101 and Inception_v3 networks, the WheNet model proved to perform better with recognition accuracies of 98.59% (Top-1%) and 99.98% (Top-5%). However, it should be noted that the use of some image preprocessing methods including image augmentation, deblurring and binarization can avoid overfitting problems and improve the performance of detection. In another study, the capacity of different configurations CNN model was studied in the classification of barley varieties comparing with image feature extracted by AlexNet and ResNet18 models based on fixed feature extractor and fine-tuning methods, and it was found that the CNN model in 64 3× 3 configurations yielded the most satisfactory result with a classification rate of 93.21% (Kozlowski et al., 2019).

Meanwhile, some other studies have also been reported on combining the CNN model with computer vision systems for detecting cereal products, for example, Cotrim et al. (2020) identified the browning degree of bread crust during baking, which was closely linked to consumer purchase decisions. In their study, the short-CNN model based on the Inception v3 module was employed to extract the image features of bread crust of seven different baking periods. The best results were obtained by short-CNN with a global accuracy of 98.8% and the short-CNN model was proven to be more advantageous than AlexNet and VGGNet-16 models.

3.3. Fruits

Fruits offer abundant essential vitamins and other nutrients and thus are important for a healthy diet. However, many factors such as harvest, storage and transportation conditions can affect their quality, therefore, the quality detection of fruits and vegetables is important for the industry.

Several studies (Hossain et al., 2019; Rodriguez et al., 2018) have illustrated that CNN has unique advantages over traditional methods in differentiating fruit varieties. In particularly, Ponce et al. (2019) combined the CNN model with machine vision systems to differentiate seven different varieties of olive fruit. They applied six different 2-D CNN architectures to extract image features of the olive fruit and found that with Inception-ResnetV2 architecture the best classification accuracy of 95.91% was achieved. Similarly, Altaheri et al. (2019) tested the capacity of pre-trained AlexNet and VGG-16 models for real-time differentiation of date fruit according to varieties, maturity, and harvesting decision, and found that the fine-tuned VGG-16 model was the best with classification accuracies of 99.01%, 97.25%, and 98.59% for varieties, maturity, and harvesting decisions, respectively.

Additionally, defect detection of fruit is a significant issue of concern. Recently, there appeared many studies about the application of CNN for defect detection of apple (Fan et al., 2020), blueberry (Qiao et al., 2020; Wang et al., 2018), winter jujube (Feng et al., 2019) and sour lemons (Jahanbakhshi et al., 2020) as well as differentiation of mealy apples from and non-mealy ones (Lashgari et al., 2020). Especially, for detecting bruises of winter jujube from four different geographical origins, Feng et al. (2019) designed a 1-D CNN model to extract pixel-wise spectral features from the region of interest (ROI) of the jujube HSI images in VIS-NIRS (502–947 nm) and NIR (975–1646 nm) spectral regions and their results proved the feasibility of 1-D CNN in detecting subtle bruises as compared with traditional methods. In addition, the feasibility of 2-D CNN combined with a computer vision system was explored by Fan et al. (2020) to detect defective apples and the best results with an accuracy of 96.5% were obtained for the testing set. Meanwhile, da Costa et al. (2020) tested the performance of different 2-D ResNet architectures based on fixed feature extractor and fine-tuning methods for detecting the external defects of tomatoes and found that fine-tuned ResNet50 yielded the most satisfactory result with classification rates of 96.2%, 94.2% and 94.6% for the training set, validation set and testing set, respectively.

More importantly, in order to improve the classification accuracy and reduce the number of architecture parameters, a deep residual 3-D CNN framework was utilized to extract simultaneously rich spectral and spatial features from hyperspectral images of fresh and decayed blueberries for the early detection of the decay and the most satisfactory result with a detection rate of 92.15% was obtained compared with AlexNet and GoogleNet models (Qiao et al., 2020). The above studies illustrated the potential of 1-D, 2-D and 3-D CNN models as an effective and rapid tool for feature extraction in defect detection of fruits.

3.4. Others

CNN has also been applied to deal with other food-related problems, including quality detection of liquid foods, food volume estimation, evaluation of nutritional contents and detection of crop diseases. Most of these studies have yielded satisfactory results, thereby encouraging further research.

3.4.1. Quality detection of liquid products

In liquid products such as milk and edible oil, adulteration can take place, thus authentication detection is vitally important to the industry. Asseiss Neto et al. (2019) designed a 1-D CNN architecture for the adulteration detection of milk. In the detection, two types of data including infrared spectra and component features were generated by Fourier transformed infrared spectroscopy (FTIR). The proposed 1-D CNN was utilized as a deep feature extractor to extract features of 1-D infrared spectra, and random forest (RF) and gradient boosting machine (GBM) classifiers were employed to analyze component features. Results showed that 1-D CNN outperformed RF, GBM and classical learning methods, with the highest classification accuracies of 98.76% and 96.95% for binary and multiclass classifications, respectively.

Recently, Wu et al. (2020) developed 2-D AlexNet architecture for

identifying three counterfeit sesame oil and four pure oil samples using 3D fluorescence spectroscopy. The proposed AlexNet was utilized as a deep feature extractor to capture the features of fluorescence contour images, which were subsequently fed into SVM and partial least squares (PLS) classifiers for determining the counterfeiting of the sesame oil. Their results revealed that the model of the AlexNet-based feature achieved better results than SVM and PLS, providing the possibility of combining CNN with fluorescence spectroscopy to rapid authentication detection of liquid products.

3.4.2. Estimation of food volume

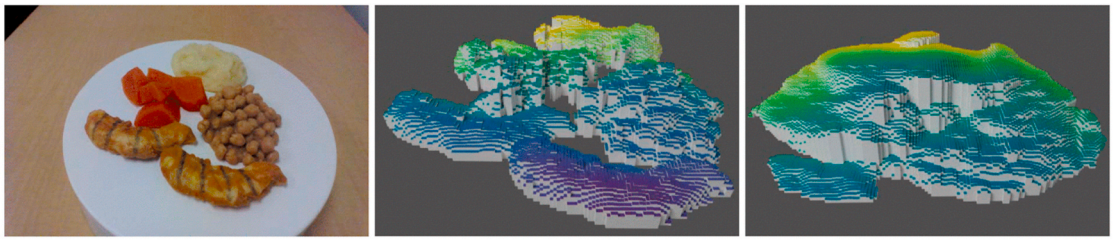
Estimation of food volume can be the most direct and effective solution to monitor dietary intake. Commonly used food volume estimation methods, such as model-based and stereo-based, have proved effective in measuring the volume of food. However, these methods rely heavily on manual intervention, which can be tedious (Lo et al., 2018). In recent years, CNN has been introduced to estimate food volume. In one study, Myers et al. (2015) designed a framework based on CNN to estimate the volume of food from a single RGB image on three food datasets including NYUv2 RGBD (training), GFood3d (fine-tuning) and NFood-3d dataset (testing), as shown in Fig. 5(a) and indicated that the proposed CNN volume predictor was very successful. Their volume assessment process could be described as follows: a CNN model was utilized to infer the depth map from a single RGB image, each pixel of the inferred depth map was then projected into a 3-D space to generate voxel representation, and the volume could thus be estimated by calculating the occupied voxels. However, their approach could generate large volume estimation errors due to view occlusion and contours ambiguity.

To address the problems of view occlusion and contours ambiguity, Lo et al. (2018) proposed a comprehensive approach based on CNN and depth-sensing technique, as shown in Fig. 5(b). The paired depth images (initial viewing angles and its opposite side) of 8 different types of foods were captured and rendered to build the training dataset based on a range of extrinsic camera parameters, which were subsequently fed into a modified neural network to train. To validate the feasibility, the trained neural network was utilized to infer the depth image of the opposite side from the input depth image with invisible viewing angles, then the two depth images were fused to generate a completed 3D point cloud map using point cloud completion algorithm. Furthermore, the iterative closest point (ICP) algorithm was adapted in the proposed approach to address the misalignment problem of point cloud registration, and then the global point cloud of the 3D model was meshed with the alpha shape to estimate the volume. Their results demonstrated that the modified neural network outperformed the naive version, and was able to infer the depth image of the opposite side from the input depth image without overfitting. However, to accurately estimate the volume from a single depth image, further investigations are required to establish a more representative 3D model database for training the neural network.

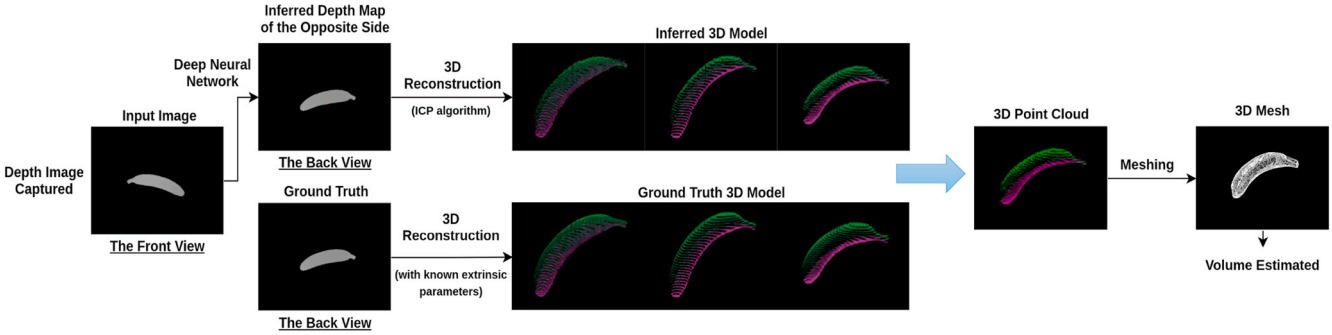
3.4.3. Evaluation of food nutritional contents

The nutrients in food that can produce calories to maintain life are protein, fat and carbohydrates. However, a high-calorie intake that is not burned by physical activity can increase the risk of developing lifestyle-related diseases such as obesity, diabetes and hypertension (Sitiju et al., 2019). In addition, high-calorie foods do not mean high nutrition. On the contrary, high-calorie foods often contain very low or no nutrients, such as some popular beverages, and junk foods. However, in addition to the nutrients mentioned above, there are mineral nutrients and vitamins. Although mineral nutrients and vitamins do not contain calories, they still have a significant relationship with our body metabolism, especially the absence of essential nutrients in the body can cause a number of health problems related to organ degradation such as weakened immune systems, weak bone structure, sparse hairline and so on (Sundaravadivel et al., 2018). Therefore, to prevent these lifestyle-related diseases, estimating the calorie intake and nutritional contents is a relevant and important issue.

Nowadays, people are concerned about high levels of calorie intakes, and evaluating the calorie from food image can provide an effective



(a) Estimate the volume of food items from voxel grid of a single RGB image using CNN



(b) Estimate the volume from the paired depth image of food items using view synthesis approach based on deep neural network

Fig. 5. The comparison of two different methods for food volume estimation using (a) the voxel grid of a depth map predicted by CNN from the RGB image of the food item (Myers et al., 2015) and (b) a 3D point cloud map generated by fusing two depth images, with one image from depth camera and the other from deep neural network (Lo et al., 2018).

solution to maintain a relatively balanced calorie content in the human body. The evaluation of food calories can be defined as a regression problem that uses multi-task 2-D CNN to predict the calories values. Therefore, Situju et al. (2019) and Ege et al. (2019) studied the feasibility of multi-task 2-D CNN architecture in order to evaluate food calorie from food images. In the study of Situju et al. (2019), two different scale image datasets were collected, including a middle-scale category-annotated food image database and a small-scale dataset containing both calorie content and salinity, which were subsequently fed into pre-trained CNNs with ImageNet to conduct two-stage fine-tuning. Their results illustrated that the multi-task CNN with two-stage fine-tuning outperformed single-task CNN and multi-task CNN without two-stage fine-tuning in terms of food classification, and calorie and salinity evaluation with a related error of 31.2%, an absolute error of 89.6 kcal and a correlation coefficient of 0.84 for calorie evaluation, and 36.1%, 0.74 g and 0.45 for salinity evaluation, respectively.

Although CNN-based methods are effective for evaluating food calorie by treating the evaluation process as a regression issue, the evaluation performance still needs improvement. Furthermore, as the calorie contents of food dishes or items in the same category can be different depending on many factors such as ingredients and cooking directions (Ege et al., 2018), it is necessary to construct large-scale food image databases labelled with calorie contents, ingredients, cooking directions and bounding boxes, respectively.

Evaluation of food nutritional contents can provide nutrition information to unpacked or cooked foods in restaurants, which is similar to the provision of nutrients in packaged food. Ahn et al. (2019) employed multiple deep neural networks (DNN) architecture to evaluate nutritional contents such as CPF (carbohydrates, proteins, and fats) of five different kinds of food items using their hyperspectral images in the wavelength range of 887–1722 nm. The proposed procedure utilized a common network to extract the common features from the hyperspectral signal from ROI, three nutrient-specific networks to compute the CPF values based on these common features, and a verification network (VN) to verify the evaluation results. Furthermore, the autoencoder mechanism and the sandwich strategy were adapted in the joint layer to share and compress the features, and the error avoidance scheme based on VN was also adapted to remove the outlier. Their results showed that the proposed procedure achieved the highest R^2 values and mean absolute percentage errors (SMAPE) of 0.9543 and 0.0997 for carbohydrate evaluation, 0.8527 and 0.1352 for protein evaluation, and 0.8481 and 0.1218 for fat evaluation.

3.4.4. Detection of crop diseases

Although not in the area of food products, crop diseases are related to food production. The efficient extraction of these diseases can enable the early detection of infected crops, avoiding the loss of food production. CNN has thus been employed for realizing efficient crop diseases detection. Li et al. (2020) and Chen et al. (2020) developed a 2-D deep convolutional neural network (DCNN) backbone and DenseNet model as an image feature extractor for the detection of paddy diseases. In the study of Li et al. (2020), a paddy diseases video detection system based on a faster-RCNN framework with a custom DCNN backbone was established. The detection process consisted of the following steps: a customized DCNN backbone was trained with still-image of three paddy diseases including sheath blight, stem borer and brown spot to extract the features of the diseases and a frame extraction module was applied to extract the frame of the input video and then the frame was sent to the faster-RCNN framework for detecting the diseases. Their results revealed that the faster-RCNN model with the customized DCNN backbone was superior to other backbone systems including VGG16, ResNet-50 and ResNet-101 and YOLOv3 models.

Additionally, detection of maize leaf disease was conducted by Priyadharshini et al. (2019), who utilized a modified 2-D LeNet architecture as a feature extractor to capture the features of three different maize disease images based on the PlantVillage dataset, and achieved the highest classification of 97.89% as compared with traditional methods of SVM and ANN. The above studies showed that integrating CNN with machine vision technique could significantly improve the detection performance of crop diseases.

4. Challenges and future work

With the unique advantages of strong feature learning and good generalization ability, CNN is potential and attractive for effective and efficient analysis of complex food matrices. CNN can not only automatically locate important features, but can also obtain unparalleled performance under challenging conditions such as complex background, and different resolutions and orientations of the images. Despite the advantages of CNN in the provision of better performance, there still remain numerous challenges to its applications in the food domain.

Firstly, many kinds of sensors and nondestructive detection techniques have been widely applied for obtaining external information including weight, smell, touch, firmness and taste and internal characteristics of food such as component contents and composition. However, the effective fusion and full utilization of multisource data by using CNN is a challenging task. Current fusion methods just simply stack and concatenate the data or features from sensors and advanced detection systems, which can bring redundant information and affect the accuracy of detection. Therefore, methods of effectively highlighting the most informative features while reducing noises are expected on data fusion.

Secondly, the robustness and generalization of the CNN model are linked to the quality and diversity of data. Although some publicly available food image datasets can be used for food recognition and classification, these datasets do not completely cover all types of foods. Moreover, some researchers use nonpublic data for food detection, which makes it even more difficult to guarantee the reproducibility of their studies. Unfortunately, building a worldwide accessible reference dataset for detection and analysis of food is challenging, the establishment of food datasets with sufficient generality and diversity is still greatly needed.

Thirdly, it remains a challenge to train CNN models and optimize the parameters according to specific food analysis tasks. The number of layers, filters and epochs, as well as the hyperparameters of the model, are usually determined by trial-and-error tuning until optimal settings are obtained, which is mostly based on expert experience and is one of the major bottlenecks of tuning for most researchers. Although recent approaches such as the automated Bayesian optimization method can find better hyperparameters faster than experts, it does not fit well with large models. Therefore, new methods should be developed so that they can not only self-adaptively search optimal settings but can also improve the performance of CNN.

Finally, CNN models deployed on mobile devices with limited memory and battery constraints bring even more challenge for real-time detection and analysis of food. In addition, the high cost of hardware needs to be taken into account.

Despite the above challenges, the enormous potentials of CNN-based approaches have not yet been fully exploited. From the perspective of improving the efficiency and accuracy of food detection, several potential research directions emerge for the CNN-based method in the future.

A promising research field is mobile hyperspectral imaging systems using snapshot HSI and 3-D CNN models. Although the snapshot HSI has the advantages in fast acquisition of spectral data and ultra-portability compared with the line-scanning HSI, it can only acquire the data in limited spectral wavelength ranges. However, due to the potential and robustness of 3-D CNN models, the snapshot HSI couple with 3-D CNN model can achieve excellent performance. Hence, the mobile HSI system is more appropriate for real-time detection of food in the future.

In addition, since HSI techniques have concepts similar to Terahertz imaging, and they are utilized as a nondestructive tool to reflect the intrinsic information of food for food inspection, it will be worth

examining the applicability of CNN coupled with Terahertz spectroscopy for food detection (Zhang et al., 2020).

Furthermore, a coming research field is combining CNN with blockchain and internet of things (IoT) techniques to control food processing as food chains usually involve a series of processes of planting (feeding), growing, harvesting (slaughtering) by farmers (butchers), processing by manufacturers, storage and transportation by distributors, and sale by retailers. IoT and blockchain can provide a large amount of information in the food chain, which can be utilized by CNN for handling and optimizing for ensuring process quality and safety. Therefore, the combination of IoT and blockchain with CNN will play an important role in food detection and analysis in the future.

5. Conclusions

CNN is a promising feature extraction tool that has gradually replaced traditional machine learning algorithms. CNN can not only extract the most robust and effective features but also has strong generalization ability, which is infeasible with traditional machine learning methods. This review introduces the principles of CNN, discusses the feature extraction methods based on 1-D, 2-D and 3-D CNN models and multi-features aggregation method, and summarizes the recent applications of CNN models in quality detection including meat and aquatic products, cereals and cereal products, fruits and vegetables, oil and milk. In addition, CNN shows to be feasible in the estimation of food volume and nutritional contents and detection of crop diseases. However, despite the prominent performance of CNN in food detection, there is still enormous potential to further exploit new algorithms for improving the computation speed of CNN, and future studies can focus on such an area. It is hoped that this review can encourage further research in food detection based on CNN.

Acknowledgements

The authors are grateful to the National Key R&D Program of China (2018YFC1603400) for its support. This research was also supported by the Guangdong Basic and Applied Basic Research Foundation (2020A1515010936), the Fundamental Research Funds for the Central Universities (D2190450), the Contemporary International Collaborative Research Centre of Guangdong Province on Food Innovative Processing and Intelligent Control (2019A050519001) and the Common Technical Innovation Team of Guangdong Province on Preservation and Logistics of Agricultural Products (2020KJ145). Yao Liu is grateful for his MSc study supervised and supported by the Academy of Contemporary Food Engineering, South China University of Technology, China.

References

Ahn, D., Choi, J.-Y., Kim, H.-C., Cho, J.-S., Moon, K.-D., & Park, T. (2019). Estimating the composition of food nutrients from hyperspectral signals based on deep neural networks. *Sensors, 19*(7), 1560.

Al-Sarayreh, M., Reis, M. M., Wei Qi, Y., & Klette, R. (2018). Detection of red-meat adulteration by deep spectral-spatial features in hyperspectral images. *Journal of Imaging, 4*(5), 63.

Al-Sarayreh, M., Reis, M. M., Yan, W. Q., & Klette, R. (2020). Potential of deep learning and snapshot hyperspectral imaging for classification of species in meat. *Food Control, 117*, 107332.

Altaheri, H., Alsulaiman, M., & Muhammad, G. (2019). Date fruit classification for robotic harvesting in a natural environment using deep learning. *IEEE Access, 7*, 117115–117133.

Asseiss Neto, H., Tavares, W. L. F., Ribeiro, D. C. S. Z., Alves, R. C. O., Fonseca, L. M., & Campos, S. V. A. (2019). On the utilization of deep and ensemble learning to detect milk adulteration. *BioData Mining, 12*(1), 1–13.

Audebert, N., Le Saux, B., & Lefevre, S. (2019). Deep learning for classification of hyperspectral data. *IEEE Geoscience and Remote Sensing Magazine, 7*(2), 159–173.

Ball, J. E., Anderson, D. T., & Chan, C. S. (2017). Comprehensive survey of deep learning in remote sensing: Theories, tools, and challenges for the community. *Journal of Applied Remote Sensing, 11*(4), Article 042609.

Chen, Y., Jiang, H., Li, C., Jia, X., & Ghamisi, P. (2016). Deep feature extraction and classification of hyperspectral images based on convolutional neural networks. *IEEE Transactions on Geoscience and Remote Sensing, 54*(10), 6232–6251.

Chen, J., Zhang, D., Nanehkaran, Y. A., & Li, D. (2020). Detection of rice plant diseases based on deep transfer learning. *Journal of the Science of Food and Agriculture, 100*(7), 3246–3256.

Cheng, W., Sun, D.-W., & Cheng, J.-H. (2016a). Pork biogenic amine index (BAI) determination based on chemometric analysis of hyperspectral imaging data. *LWT-Food Science and Technology, 73*, 13–19.

Cheng, W., Sun, D.-W., Pu, H., & Liu, Y. (2016b). Integration of spectral and textural data for enhancing hyperspectral prediction of K value in pork meat. *LWT-Food Science and Technology, 72*, 322–329.

Cheng, W., Sun, D.-W., Pu, H., & Wei, Q. (2017). Chemical spoilage extent traceability of two kinds of processed pork meats using one multispectral system developed by hyperspectral imaging combined with effective variable selection methods. *Food Chemistry, 221*, 1989–1996.

Cheng, W., Sun, D.-W., Pu, H., & Wei, Q. (2018). Heterospectral two-dimensional correlation analysis with near-infrared hyperspectral imaging for monitoring oxidative damage of pork myofibrils during frozen storage. *Food Chemistry, 248*, 119–127.

Ciocca, G., Napoletano, P., & Schettini, R. (2018). CNN-based features for retrieval and classification of food images. *Computer Vision and Image Understanding, 176*, 70–77.

Cotrim, W.d. S., Rodridgues Minim, V. P., Felix, L. B., & Minim, L. A. (2020). Short convolutional neural networks applied to the recognition of the browning stages of bread crust. *Journal of Food Engineering, 277*, 109916, 109916.

da Costa, A. Z., Figueroa, H. E. H., & Fracarolli, J. A. (2020). Computer vision based detection of external defects on tomatoes using deep learning. *Biosystems Engineering, 190*, 131–144.

Dai, Q., Cheng, J.-H., Sun, D.-W., Zhu, Z., & Pu, H. (2016). Prediction of total volatile basic nitrogen contents using wavelet features from visible/near-infrared hyperspectral images of prawn (Metapenaeus ensis). *Food Chemistry, 197*, 257–265.

Du, C. J., & Sun, D.-W. (2005). Pizza sauce spread classification using colour vision and support vector machines. *Journal of Food Engineering, 66*(2), 137–145.

Ege, T., & Yanai, K. (2018). Image-based food calorie estimation using recipe information. *IEICE - Transactions on Info and Systems, E101D*(5), 1333–1341.

Ege, T., & Yanai, K. (2019). Simultaneous estimation of dish locations and calories with multi-task learning. *IEICE - Transactions on Info and Systems, E102D*(7), 1240–1246.

Elmasry, G., Barbin, D. F., Sun, D.-W., & Allen, P. (2012). Meat quality evaluation by hyperspectral imaging technique: An overview. *Critical Reviews in Food Science and Nutrition, 52*(8), 689–711.

Fan, S., Li, J., Zhang, Y., Tian, X., Wang, Q., He, X., Zhang, C., & Huang, W. (2020). On line detection of defective apples using computer vision system combined with deep learning methods. *Journal of Food Engineering, 286*, 110102.

Feng, L., Zhu, S., Zhou, L., Zhao, Y., Bao, Y., Zhang, C., & He, Y. (2019). Detection of subtle bruises on winter jujube using hyperspectral imaging with pixel-wise deep learning method. *IEEE Access, 7*, 64494–64505.

Gu, K., Xia, Z., Qiao, J., & Lin, W. (2020). Deep dual-channel neural network for image-based smoke detection. *IEEE Transactions on Multimedia, 22*(2), 311–323.

Gu, K., Zhang, Y., & Qiao, J. (2021). Ensemble meta-learning for few-shot soot density recognition. *IEEE Transactions on Industrial Informatics, 17*(3), 2261–2270.

Hossain, M. S., Al-Hammadi, M., & Muhammad, G. (2019). Automatic fruit classification using deep learning for industrial applications. *IEEE Transactions on Industrial Informatics, 15*(2), 1027–1034.

Hu, W., Huang, Y., Wei, L., Zhang, F., & Li, H. (2015). Deep convolutional neural networks for hyperspectral image classification. *Journal of Sensors*, 258619, 2015.

Hussain, G., Maheshwari, M. K., Memon, M. L., Jabbar, M. S., & Javed, K. (2019). A CNN based automated activity and food recognition using wearable sensor for preventive healthcare. *Electronics, 8*(12), 1425.

Jackman, P., Sun, D.-W., & Allen, P. (2011). Recent advances in the use of computer vision technology in the quality assessment of fresh meats. *Trends in Food Science & Technology, 22*(4), 185–197.

Jahanbakhshi, A., Momeny, M., Mahmoudi, M., & Zhang, Y.-D. (2020). Classification of sour lemons based on apparent defects using stochastic pooling mechanism in deep convolutional neural networks. *Scientia Horticulturae, 263*, 109133.

Jahani Heravi, E., Habibi Aghdam, H., & Puig, D. (2018). An optimized convolutional neural network with bottleneck and spatial pyramid pooling layers for classification of foods. *Pattern Recognition Letters, 105*, 50–58.

Jiang, S., Min, W., Liu, L., & Luo, Z. (2020). Multi-scale multi-view deep feature aggregation for food recognition. *IEEE Transactions on Image Processing, 29*(1), 265–276.

Kamilaris, A., & Prenafeta-Boldu, F. X. (2018a). Deep learning in agriculture: A survey. *Computers and Electronics in Agriculture, 147*, 70–90.

Kamilaris, A., & Prenafeta-Boldu, F. X. (2018b). A review of the use of convolutional neural networks in agriculture. *Journal of Agricultural Science, 156*(3), 312–322.

Kozlowski, M., Gorecki, P., & Szczypinski, P. M. (2019). Varietal classification of barley by convolutional neural networks. *Biosystems Engineering, 184*, 155–165.

Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2012). ImageNet classification with deep convolutional neural networks. *Communications of the ACM, 60*(6), 84–90.

Kumar, B., Dikshit, O., Gupta, A., & Singh, M. K. (2020). Feature extraction for hyperspectral image classification: A review. *International Journal of Remote Sensing, 41*(16), 6248–6287.

Lashgari, M., Imanmehr, A., & Tavakoli, H. (2020). Fusion of acoustic sensing and deep learning techniques for apple mealiness detection. *Journal of Food Science and Technology-Mysore, 57*(6), 2233–2240.

LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. *Nature, 521*(7553), 436–444.

Lecun, Y., Bottou, L., Bengio, Y., & Haffner, P. (1998). Gradient-based learning applied to document recognition. *Proceedings of the IEEE, 86*(11), 2278–2324.

Lin, P., Li, X. L., Chen, Y. M., & He, Y. (2018). A deep convolutional neural network architecture for boosting image discrimination accuracy of rice species. *Food and Bioprocess Technology, 11*(4), 765–773.
Liu, C., Cao, Y., Luo, Y., Chen, G., Vokkarane, V., Ma, Y., … Hou, P. (2018a). A new deep learning-based food recognition system for dietary assessment on an edge computing service infrastructure. *IEEE Transactions on Services Computing, 11*(2), 249–261.
Liu, Y., Sun, D.-W., Cheng, J.-H., & Han, Z. (2018b). Hyperspectral imaging sensing of changes in moisture content and color of beef during microwave heating process. *Food Analytical Methods, 11*(9), 2472–2484.
Liu, Y., Pu, H., & Sun, D.-W. (2017). Hyperspectral imaging technique for evaluating food quality and safety during various processes: A review of recent applications. *Trends in Food Science & Technology, 69*, 25–35.
Li, D., Wang, R., Xie, C., Liu, L., Zhang, J., Li, R., Wang, F., Zhou, M., & Liu, W. (2020). A recognition method for rice plant diseases and pests video detection based on deep convolutional neural network. *Sensors, 20*(3), 578.
Li, Y., Zhang, H., & Shen, Q. (2017). Spectral-spatial classification of hyperspectral imagery with 3D convolutional neural network. *Remote Sensing, 9*(1), 67.
Lo, F. P. W., Sun, Y., Qiu, J., & Lo, B. (2018). Food volume estimation based on deep learning view synthesis from a single depth map. *Nutrients, 10*(12), 2005.
Ma, J., Pu, H., & Sun, D.-W. (2018). Predicting intramuscular fat content variations in boiled pork muscles by hyperspectral imaging using a novel spectral pre-processing technique. *LWT-Food Science and Technology, 94*, 119–128.
Ma, J., Sun, D.-W., & Pu, H. (2017). Model improvement for predicting moisture content (MC) in pork longissimus dorsi muscles under diverse processing conditions by hyperspectral imaging. *Journal of Food Engineering, 196*, 65–72.
McAllister, P., Zheng, H., Bond, R., & Moorhead, A. (2018). Combining deep residual neural network features with supervised machine learning algorithms to classify diverse food image datasets. *Computers in Biology and Medicine, 95*, 217–233.
Mezgec, S., & Seljak, B. K. (2017). NutriNet: A deep learning food and drink image recognition system for dietary assessment. *Nutrients, 9*(7), 657.
Myers, A., Johnston, N., Rathod, V., Korattikara, A., Gorban, A., Silberman, N., Guadarrama, S., Papandreou, G., Huang, J., & Murphy, K. (2015). Im2Calories: Towards an automated mobile vision food diary. In *International conference on computer vision (ICCV), 2015 IEEE* (pp. 1233–1241). IEEE.
Ni, C., Wang, D., Vinson, R., Holmes, M., & Tao, Y. (2019). Automatic inspection machine for maize kernels based on deep convolutional neural networks. *Biosystems Engineering, 178*, 131–144.
Pandey, P., Deepthi, A., Mandal, B., & Puhan, N. B. (2017). FoodNet: Recognizing foods using ensemble of deep networks. *IEEE Signal Processing Letters, 24*(12), 1758–1762.
Pan, L., Li, C., Pouyanfar, S., Chen, R., & Zhou, Y. (2020). A novel combinational convolutional neural network for automatic food-ingredient classification. *Cmc-Computers Materials & Continua, 62*(2), 731–746.
Pan, L., Qin, J., Chen, H., Xiang, X., Li, C., & Chen, R. (2019). Image augmentation-based food recognition with convolutional neural networks. *Cmc-Computers Materials & Continua, 59*(1), 297–313.
Pan, Y., Sun, D.-W., Cheng, J.-H., & Han, Z. (2018). Non-destructive detection and screening of non-uniformity in microwave sterilization using hyperspectral imaging analysis. *Food Analytical Methods, 11*(6), 1568–1580.
Paoletti, M. E., Haut, J. M., Plaza, J., & Plaza, A. (2019). Deep learning classifiers for hyperspectral imaging: A review. *ISPRS Journal of Photogrammetry and Remote Sensing, 158*, 279–317.
Ponce, J. M., Aquino, A., & Andujar, J. M. (2019). Olive-fruit variety classification by means of image processing and convolutional neural networks. *IEEE Access, 7*, 147629–147641.
Pouladzadeh, P., & Shirmohammadi, S. (2017). Mobile multi-food recognition using deep learning. *ACM Transactions on Multimedia Computing, Communications, and Applications, 13*(3), 36.
Priyadharshini, R. A., Arivazhagan, S., Arun, M., & Mirnalini, A. (2019). Maize leaf disease classification using deep convolutional neural networks. *Neural Computing & Applications, 31*(12), 8887–8895.
Qiao, S., Wang, Q., Zhang, J., & Pei, Z. (2020). Detection and classification of early decay on blueberry based on improved deep residual 3D convolutional neural network in hyperspectral images. *Scientific Programming, 2020*, Article 8895875.
Qiu, Z., Chen, J., Zhao, Y., Zhu, S., He, Y., & Zhang, C. (2018). Variety identification of single rice seed using hyperspectral imaging combined with convolutional neural network. *Applied Sciences-Basel, 8*(2), 212.
Qi, W., Zhang, X., Wang, N., Zhang, M., & Cen, Y. (2019). A spectral-spatial cascaded 3D convolutional neural network with a convolutional long short-term memory network for hyperspectral image classification. *Remote Sensing, 11*(20), 2363.
Qu, J.-H., Liu, D., Cheng, J.-H., Sun, D.-W., Ma, J., Pu, H., & Zeng, X.-A. (2015). Applications of near-infrared spectroscopy in food safety evaluation and control: A review of recent research advances. *Critical Reviews in Food Science and Nutrition, 55*(13), 1939–1954.
Rodriguez, F. J., Garcia, A., Pardo, P. J., Chavez, F., & Luque-Baena, R. M. (2018). Study and classification of plum varieties using image analysis and deep learning techniques. *Progress in Artificial Intelligence, 7*(2), 119–127.
Shen, Y., Yin, Y., Zhao, C., Li, B., Wang, J., Li, G., & Zhang, Z. (2019). Image recognition method based on an improved convolutional neural network to detect impurities in wheat. *IEEE Access, 7*, 162206–162218.
Situju, S. F., Takimoto, H., Sato, S., Yamauchi, H., Kanagawa, A., & Lawi, A. (2019). Food constituent estimation for lifestyle disease prevention by multi-task CNN. *Applied Artificial Intelligence, 33*(8), 732–746.
Steinbrener, J., Posch, K., & Leitner, R. (2019). Hyperspectral fruit and vegetable classification using convolutional neural networks. *Computers and Electronics in Agriculture, 162*, 364–372.
Sun, D.-W., & Brosnan, T. (2003). Pizza quality evaluation using computer vision - Part 2 - Pizza topping analysis. *Journal of Food Engineering, 57*(1), 91–95.
Sundaravadivel, P., Kesavan, K., Kesavan, O., Mohanty, S. P., & Kougianos, E. (2018). Smart-log: A deep-learning based automated nutrition monitoring system in the IoT. *IEEE Transactions on Consumer Electronics, 64*(3), 390–398.
Taheri-Garavand, A., Nasiri, A., Banan, A., & Zhang, Y.-D. (2020). Smart deep learning-based approach for non-destructive freshness diagnosis of common carp fish. *Journal of Food Engineering, 278*, 109930.
Teng, J., Zhang, D., Lee, D.-J., & Chou, Y. (2019). Recognition of Chinese food using convolutional neural network. *Multimedia Tools and Applications, 78*(9), 11155–11172.
Wang, Z., Hu, M., & Zhai, G. (2018). Application of deep learning architectures for accurate and rapid detection of internal mechanical damage of blueberry using hyperspectral transmittance data. *Sensors, 18*(4), 1126.
Wang, H. H., & Sun, D.-W. (2003). Assessment of cheese browning affected by baking conditions using computer vision. *Journal of Food Engineering, 56*(4), 339–345.
Wang, K., Sun, D.-W., & Pu, H. (2017). Emerging non-destructive terahertz spectroscopic imaging technique: Principle and applications in the agri-food industry. *Trends in Food Science & Technology, 67*, 93–105.
Wu, N., Zhang, C., Bai, X., Du, X., & He, Y. (2018). Discrimination of chrysanthemum varieties using hyperspectral imaging combined with a deep convolutional neural network. *Molecules, 23*(11), 2831.
Wu, X., Zhao, Z., Tian, R., Shang, Z., & Liu, H. (2020). Identification and quantification of counterfeit sesame oil by 3D fluorescence spectroscopy and convolutional neural network. *Food Chemistry, 311*, 125882.
Xu, J.-L., & Sun, D.-W. (2018). Computer vision detection of salmon muscle gaping using convolutional neural network features. *Food Analytical Methods, 11*(1), 34–47.
Zhang, J., Yang, Y., Feng, X., Xu, H., Chen, J., & He, Y. (2020). Identification of bacterial blight resistant rice seeds using terahertz imaging and hyperspectral imaging combined with convolutional neural network. *Frontiers of Plant Science, 11*, 821.
Zheng, C., Sun, D.-W., & Zheng, L. (2006). Correlating colour to moisture content of large cooked beef joints by computer vision. *Journal of Food Engineering, 77*(4), 858–863.
Zhou, L., Zhang, C., Liu, F., Qiu, Z., & He, Y. (2019). Application of deep learning in food: A review. *Comprehensive Reviews in Food Science and Food Safety, 18*(6), 1793–1811.
Zhu, S., Zhou, L., Zhang, C., Bao, Y., Wu, B., Chu, H., Yu, Y., He, Y., & Feng, L. (2019). Identification of soybean varieties using hyperspectral imaging coupled with convolutional neural network. *Sensors, 19*(19), 4065.