Application of Deep Learning in Food: A Review

Lei Zhou, Chu Zhang, Fei Liu, Zhengjun Qiu (iD), and Yong He

Abstract: Deep learning has been proved to be an advanced technology for big data analysis with a large number of successful cases in image processing, speech recognition, object detection, and so on. Recently, it has also been introduced in food science and engineering. To our knowledge, this review is the first in the food domain. In this paper, we provided a brief introduction of deep learning and detailedly described the structure of some popular architectures of deep neural networks and the approaches for training a model. We surveyed dozens of articles that used deep learning as the data analysis tool to solve the problems and challenges in food domain, including food recognition, calories estimation, quality detection of fruits, vegetables, meat and aquatic products, food supply chain, and food contamination. The specific problems, the datasets, the preprocessing methods, the networks and frameworks used, the performance achieved, and the comparison with other popular solutions of each research were investigated. We also analyzed the potential of deep learning to be used as an advanced data mining tool in food sensory and consume researches. The result of our survey indicates that deep learning outperforms other methods such as manual feature extractors, conventional machine learning algorithms, and deep learning as a promising tool in food quality and safety inspection. The encouraging results in classification and regression problems achieved by deep learning will attract more research efforts to apply deep learning into the field of food in the future.

Keywords: computer vision, deep learning, food quality, food recognition, spectroscopy

Introduction

Healthy diet is important to human health (de Ridder, Kroese, Evers, Adriaanse, & Gillebaart, 2017). Natural products have been widely used as food, and they can also be processed to meet the demand of consumers. Food (natural products and processed food) attributes such as type, compositions, nutrients, and process styles are concerned issues for healthy diet. It is a fact that people from different regions have different eating habits. Knowing the attributes of food (type, compositions, nutrients and process styles, and so on) is important to inspect food quality and safety for consumers all over the world (Lule & Xia, 2005).

Rapid, accurate, and automatic determination of food attributes is a practical demand in daily life. Modern techniques, including electronic noses (Tian, Li, Qin, Yu, & Ma, 2014), computer vision (Brosnan & Sun, 2004), spectroscopy and spectral imaging (Barbin, Felicio, Sun, Nixdorf, & Hirooka, 2014), and so on, have been widely used to detect food attributes. These techniques can acquire a large amount of digital information relating to food properties. Data analysis of these techniques is extremely important due to the fact that the large amount data contain much redundant and irrelevant information. How to deal with such a large amount of data and extract useful features from the acquired data is an urgent

and important issue, and also a challenge to bring these techniques into real-world application (APP).

Many data analysis methods have been developed to deal with the large amount of data, for modeling such as partial least squares (PLS) (J.H. Cheng & Sun, 2017), artificial neural network (ANN) (Yiqun, Kangas, & Rasco, 2007), support vector machine (SVM) (Pouladzadeh, Villalobos, Almaghrabi, & Shirmohammadi, 2012), random forest (Bossard, Guillaumin, & Gool, 2014), k-nearest neighbor (KNN) (Yordi et al., 2015), and so on. For feature extraction, such as principal component analysis (PCA) (Granato, Santos, Escher, Ferreira, & Maggio, 2018), wavelet transform (WT) (Ma, 2017), independent component correlation algorithm (ICA) (Monakhova, Tsikin, Kuballa, Lachenmeier, & Mushtakova, 2014), scale-invariant feature transform (Giovany, Putra, Hariawan, & Wulandhari, 2017), speedup robust features (Bay, Ess, Tuytelaars, & Van Gool, 2008), histogram of oriented gradient (Ahmed & Ozeki, 2015), and so on. These methods have showed their great value in dealing with these data.

Deep learning, as an effective machine learning algorithm, has been widely studied (LeCun, Bengio, & Hinton, 2015) and now attracts more attentions from various fields such as remote sensing (G. Cheng & Han, 2016), agriculture production (Kamilaris & Prenafeta-Boldu, 2018), medical science (Shen, Wu, & Suk, 2017), robotics (Pierson & Gashler, 2017), healthcare (Miotto, Wang, Wang, Jiang, & Dudley, 2018), human action recognition (D. Wu, Sharma, & Blumenstein, 2017), speech recognition (Noda, Yamaguchi, Nakadai, Okuno, & Ogata, 2015), and so on. Deep learning has showed significant advantages in automatically learning data representations (even for multidomain feature
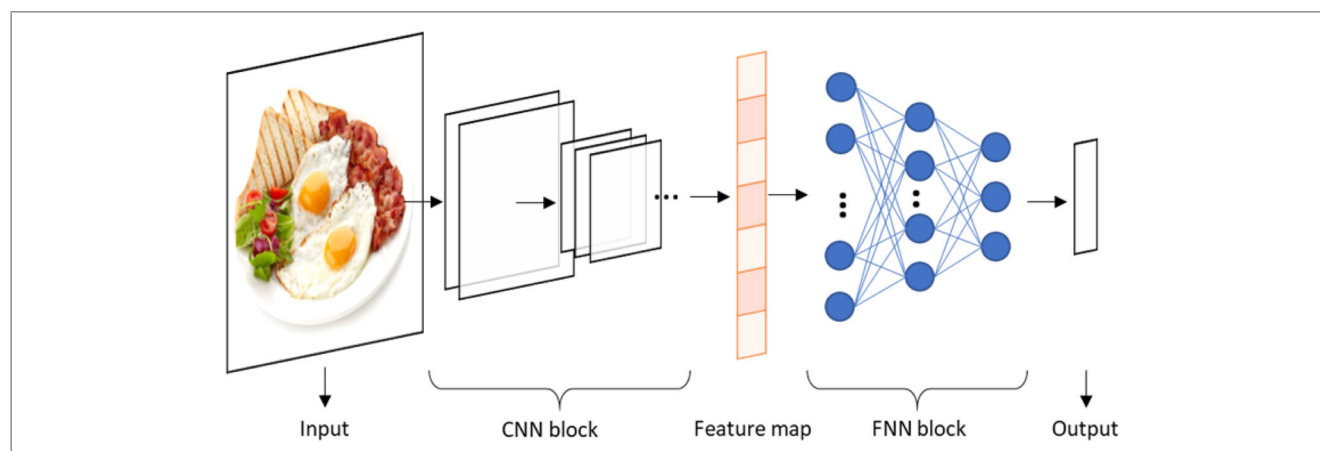
Figure 1–A typical CNN structure for image classification. A set of quadrilaterals stacked together represent the output after the calculation by convolutional layers. The circular symbols filled with blue color in this figure represent neurons in FNN block, and the lines between circular symbols denote the weights and bias. The black arrows between different layers denote the direction of data transmission.

extraction), transfer learning (Ng, Nguyen, Vonikakis, & Winkler, 2015), dealing with the large amount of data, and obtaining better performance and higher precision (Kamilaris & Prenafeta-Boldu, 2018). Convolutional neural network (CNN) and its derivative algorithms have been recognized as the key methods in most of the surveyed articles, which can automatically learn deep features of input digital information for subsequent classification or regression tasks. The large amount of data collected by the tools for food quality and safety evaluation (spectroscopy, electronic nose, digital cameras, and so on) can be successfully processed by CNN. It is worth mentioning that CNN has been proved to be effective in image analysis (two-dimensional data) (Krizhevsky, Sutskever, & Hinton, 2012), and has been extended to one-dimensional (N. Wu, Zhang, Bai, Du, & He, 2018) and three-dimensional data (Jun et al., 2018) to handle more diverse data formats.

Nowadays, deep learning has been introduced into food field by analyzing RGB images (Kawano & Yanai, 2014) and spectra images (Al-Sarayreh, Reis, Yan, & Klette, 2018) of food. However, due to the fact that understanding and APP of deep learning is a difficult issue for researchers and workers related in food industry, the researchers are on the way. The objective of this survey is to present a comprehensive overview of the latest research progress in the APP of deep learning in food field, and provide guidance for researchers and workers in this field.

Brief introduction of deep learning

Machine learning has been active in various fields, which acts as an effective tool for data processing. For the lack of ability to analyze raw natural data, traditional machine learning techniques usually needs to be supplemented by a manual feature extraction method. With the development of hardware computing ability and storage capacity, the abilities of machine learning can be promoted by adding more complex structures to achieve deep representation of the data (Schmidhuber, 2015). Representation learning enables a machine to extract the features from raw data for detection, classification, or regression. Deep learning can be understood as a kind of representation-learning method that refines multilevel representation by utilizing the deep ANN composed by multiple layers of neurons (nonlinear modules). Due to the strong feature learning ability of deep learning method, many complex problems can be solved in a rapid and effective way. Deep learning models demonstrate powerful capabilities in classification/regression tasks,

provided that adequate data support was available which represents the specific problem. With the strong ability of automatic feature learning, deep learning method starts to be applied in the field of food science, mainly referring to food category recognition, fruit and vegetable quality detection, food calorie estimation, and so on. We will introduce in detail in section "Deep learning applications in food."

CNN including a set of components (convolutional layers, pooling layers, fully connected layers, and so on) is currently considered as one of the most popular machine intelligence models for big data analysis in various research areas. A typical architecture of CNN model for classification problems is displayed in Figure 1. Convolution operations are implemented by traversing input matrices with convolution kernels that can be understood as filters for feature extraction. Different from filters used in conventional image processing method whose parameters need to be set manually, the parameters inside the kernel can be learned automatically by deep learning method. Convolutional layers are built by a set of convolution kernels, whose parameters (channels, kernel size, strides, padding, activation, and so on) should be set and optimized according to the practical problem. The computed output from convolutional layer is then subsampled by pooling layers. A group of chained convolutional layers and pooling layers can learn high-level features representing the original input. The fully connected network (FNN) block, composed by fully connected neural units, is usually placed at the end as the classifier or used to generate numerical output for regression problems exploiting the learned feature map.

Figure 2 illustrates another widely used deep learning model named stacked autoencoders (SAEs) (Suk, Lee, Shen, & Alzheimer's Dis, 2015), whose structure is similar to conventional ANNs. SAE can be used as an unsupervised learning method to obtain features from input. See Figure 2(A), the input is encoded into a vector with fewer dimensions than that of itself. Then, the vector is extended by the decoder to reconstruct the original input. Thus, this vector can be used as the deep features of the raw input. The trained encoder coupled with FNN shown in Figure 2(B) can be employed as a deep learning model to solve classification or regression problems.

Besides classification and regression, deep learning also demonstrates strong capabilities to process image segmentation tasks. Fully convolutional network (FCN) (Shelhamer, Long, & Darrell, 2017)
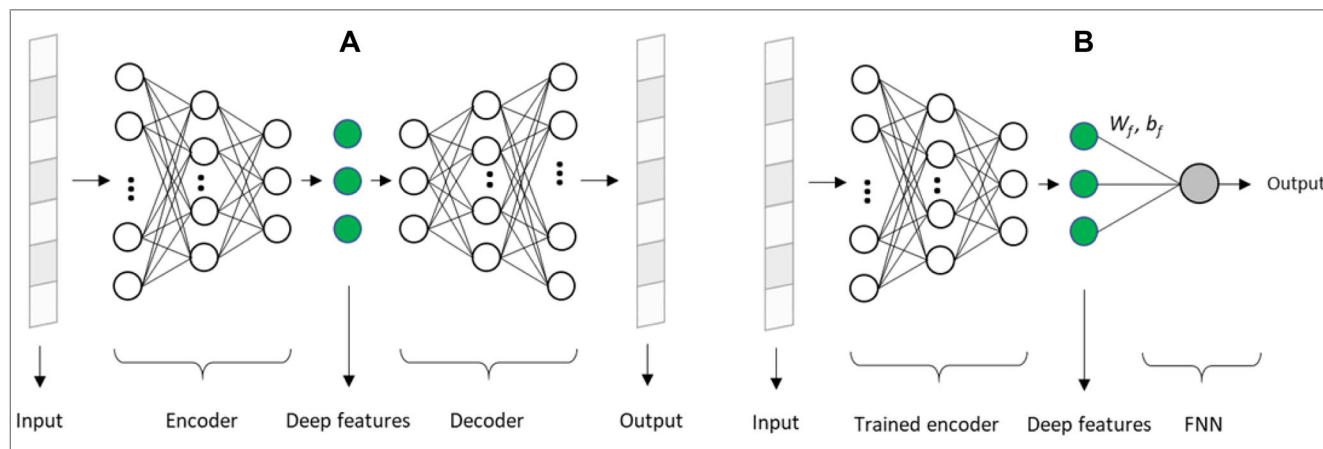
Figure 2–Typical SAE structure. (A) SAE model composed of an encoder and a decoder that would be trained. (B) The decoder is removed and the encoder is retained, followed by a regression network for prediction. The circular symbols colored green denote the neurons in feature layer that could output the deep features learned by SAE model, and W_f and b_f denote the weights and bias in FNN block.
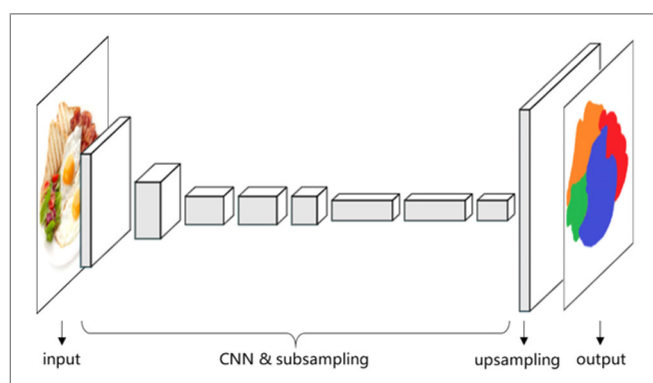

Figure 3–Structure of FCN. The cubes of different size denote the shapes (rows, columns, and channels) of matrixes output from the different convolutional layers. The areas marked with different colors of the output represent different categories of ingredients.

is a landmark progress in the field of image segmentation. Similar to general convolution networks, FCN contains several convolution layers for feature extraction, but the FNN block in an ordinary CNN is replaced by a deconvolution layer that upsamples the feature map and expands the same width and height of output as those of the input images. For instance, when a food image is fed into an FCN model, the CNN and subsampling block extracts valuable information of the input food image. Then, the upsampling block outputs the segmentation result that is an image with the same rows and columns as the input. Each pixel inside the output image represents a category. In Figure 3, red, blue, green, and orange pixels represent "beef," "eggs," "vegetables," and "bread," respectively. Thus, FCN can be regarded as a pixel-to-pixel network.

The following is an explanation to the training and optimization process of the network: The weights of deep neural networks (DNNs) are initialized randomly or by Xavier (Glorot & Bengio, 2010) method and tuned during the training procedure including forward and back propagation process. In the forward propagation process, the difference between the output value (or predicted value) and the label value (or ground truth) is calculated according to the defined loss function. In the backpropagation process, the weight of the neural network is updated to minimize the loss function via Stochastic Gradient Descent (SGD) (Ketkar, 2014),

AdaGrad (Duchi, Hazan, & Singer, 2011), and Adam (Kingma & Ba, 2014) algorithms. The hyperparameters of the network, such as learning (which controls the pace of weight adjustment), batch size, number of convolution kernels and layers, and so on, could be fitted by evaluating the performance (the output of loss function) on validation set. Besides, the combination of feature graphs, the form of convolution kernels, and the parallel network structure should be considered depending on the specific problems.

Besides the three kinds of frequently used models mentioned in this section, there are many other network structures such as recurrent neural network (RNN) (Graves, Mohamed, & Hinton, 2013), long short-term memory (LSTM) (Hochreiter & Schmidhuber, 1997), single-shot multibox detection (SSD) (Liu et al., 2016b), region-based CNN) (Ren, He, Girshick, & Sun, 2017), sequence to sequence (Sutskever, Vinyals, & Le, 2014), textCNN (Kim, 2014), and so on. Processing data types are not only limited to RGB images, but also can be any other forms of data such as video, audio, voice, text, natural language, and so on (Kamilaris & Prenafeta-Boldu, 2018).

Deep Learning Applications in Food
Food recognition and classification

The diets and eating habits can affect health of human beings. Especially for diabetics, allergic people, and so on, they should strictly monitor and control their dietary behavior. Food recognition and classification is an important task to help human beings record the daily diets. Images of food are one of the most important information to reflect the characteristics of food. Moreover, image sensing is a relatively easy and low-cost information acquisition tool for food appearance analysis. For natural products like food and processed food, the large variations in food shape, volume, texture, color, and compositions make food recognition a challenging task. Various background and layout of food stuffs also introduce variations for food recognition and classification. At present, due to the common use of CNN, image analysis has been the most commonly used pattern in food recognition and classification.

There are various popular CNN architectures for image processing including AlexNet (Krizhevsky et al., 2012), a network using repetitive units called visual geometry group network (VGG) (Simonyan & Zisserman, 2014), GoogLeNet (Szegedy et al., 2015) that includes parallel data channels, and residual neural network (ResNet) (He, Zhang, Ren, & Sun, 2016) constructed by residual

Table 1–Performances of CNN-based approaches achieved on three benchmarked databases.

Database	Model and reference	Top-1%	Top-5%
Food-101: 101 food categories, with 1000 images per category	Author defined (Heravi, Aghdam, & Puig, 2018)	65.40	87.00
	CNN-FOOD (Yanai & Kawano, 2015)	70.41	/
	Multitask (H. Wu, Merler, Uceda-Sosa, & Smith, 2016)	72.11	/
	FoodNet (Pandey, Deepthi, Mandal, & Puhan, 2017)	72.12	91.61
	DeepFood (Liu et al., 2016a)	77.40	93.70
	Inception Module (C. Liu et al., 2018)	77.00	94.00
	ResNet (Fu et al., 2017)	78.50	94.10
	ResNet-50 (Ciocca, Napoletano, & Schettini, 2018)	82.54	95.79
	Inception-v3+FPCNN (Zheng, Zou, & Wang, 2018)	87.96	/
	Inception V3 (Hassannejad et al., 2016)	88.28	96.88
	wide-slice residual networks (WISeR) (Martinel, Foresti, & Micheloni, 2018)	90.27	98.71
UECFood-256: 256 Japanese food image categories	DeepFood (Liu et al., 2016a)	54.70	81.50
	Author defined (Heravi et al., 2018)	60.00	85.00
	Inception Module (C. Liu et al., 2018)	63.60	87.00
	DeepFoodCam (Kawano & Yanai, 2014)	63.77	85.82
	CNN-FOOD (Yanai & Kawano, 2015)	67.57	88.97
	ResNet (Fu et al., 2017)	71.20	91.10
	ResNet-50 (Ciocca et al., 2018)	71.70	91.33
	Inception V3 (Hassannejad et al., 2016)	76.17	92.58
	Inception-v3+FP-CNN (Zheng et al., 2018)	78.60	/
	WISeR (Martinel et al., 2018)	83.15	95.45
UECFood-100: Images of 100 kinds of Japanese food with at least 100 images per category	DeepFoodCam (Kawano & Yanai, 2014)	72.26	92.00
	DeepFood (Liu et al., 2016a)	76.3	94.60
	Inception Module (C. Liu et al., 2018)	76.30	94.60
	CNN-FOOD (Yanai & Kawano, 2015)	78.48	94.85
	ResNet (Fu et al., 2017)	80.60	95.90
	Inception V3 (Hassannejad et al., 2016)	81.45	97.27
	MultiTaskCNN (Chen & Ngo, 2016)	82.12	97.29
	Inception-v3+FP-CNN (Zheng et al., 2018)	86.51	/
	WISeR (Martinel et al., 2018)	89.58	99.23

blocks. Furthermore, these mentioned network architectures can be downloaded from the model zoo with pretrained weights. That is, the models have already been trained by some image datasets like ImageNet (Deng et al., 2009) so that these pretraining models have already learned the ability to extract image features (such as colors, texture information, high-level abstract representations, and so on). Researches can use their specific image datasets to implement transfer learning based on the pretrained model, which means that we can use our dataset to retrain the weights of fully connected structure for final classification while keeping the weights of convolution layer unchanged, or slightly adjusting the weights of the whole network. The mentioned retraining method is called "fine-tuning," which is proved to be a successful method to shorten the training time and to gain a more accurate result. Convolutional networks have been widely used in food/nonfood classification, food category discrimination, and ingredients identification.

Food compared with nonfood discrimination is a binary classification issue that can be effectively solved via deep learning. Singla, Yuan, and Ebrahimi (2016) created a database named Food-5K (available at http://mmspg.ep.ch/food-image-datasets) consisting of 2500 food images (selected from three popular image sets for food recognition: Food-101, UECFood-100 and UECFood-256) and 2500 images of other objects. A fine-tuned GoogLeNet model was deployed and a satisfactory result was achieved with the accuracy of 99.2%. W. Jia et al. (2018) reported an accuracy of 98.7% on Food-5K database for binary classification, and McAllister, Zheng, Bond, and Moorhead (2018) reached the highest accuracy on Food-5K (99.4% for validation dataset and 98.8% for evaluation dataset) using Radial Basis Function (RBF) kernel-based SVM with ResNet-152. Ragusa, Tomaselli, Furnari, Battiato, and Farinella (2016) achieved 94.86% for classification accuracy on another database (8005 nonfood images and 3583 food images from Flickr and 3583 food from UNICT-FD889) by coupling fine-tuned AlexNet with a binary SVM classifier.

After recognizing the images that contain food, further work to study is food classification, which is a multiclassification problem. There existed several open-access food image datasets with different categories such as Food-101 (Bossard, Guillaumin, & Gool, 2014), UECFood-256 (Kawano & Yanai, 2015), UECFood-100 (Matsuda, Hoashi, & Yanai, 2012), and so on, which are summarized as Table 1. These large food images sets are possible to provide adequate features of food image for training a DNN model for food recognition. In the surveyed researches, most of them used these common datasets to train a classifier and evaluate the trained model, while the others supplemented their own collected images based on public datasets and performed the experiments.

Food-101 database including 101 food classes with 1000 image of each class is a popular dataset in food domain. Bossard et al. (2014) created the Food-101 database and achieved an average accuracy of 50.76% for classification using traditional machine learning methods. Later, many deep learning-based classifiers for food were trained using this database. The most concerned evaluation indicators in food classification tasks are Top-1 classification accuracy (Top-1%) and Top-5 classification accuracy (Top-5%). As for the former, the predicted label takes the largest one in the output probability vector as the predicted result. If the classification of the most probable one in the predicted result is correct, the predicted result is correct. Otherwise, the prediction is wrong. As for the latter, it takes the top five of the largest probability vectors, and the prediction is correct as long as the correct probability occurs. Otherwise, the prediction is wrong. Tatsuma and Aono (2016) reported a new approach for food classification by using the covariances of features of trained CNN as the representation of images, and achieved 58.65% for average accuracy. Yanai and Kawano (2015) used the fine-tuned AlexNet to achieve the top-1 accuracy for 70.41%. Liu et al. (2016a) presented a network named Deep Food reached 77.40% and 93.70% for Top-1 and Top-5%, respectively. Fu, Chen, and Li (2017) obtained a better result of
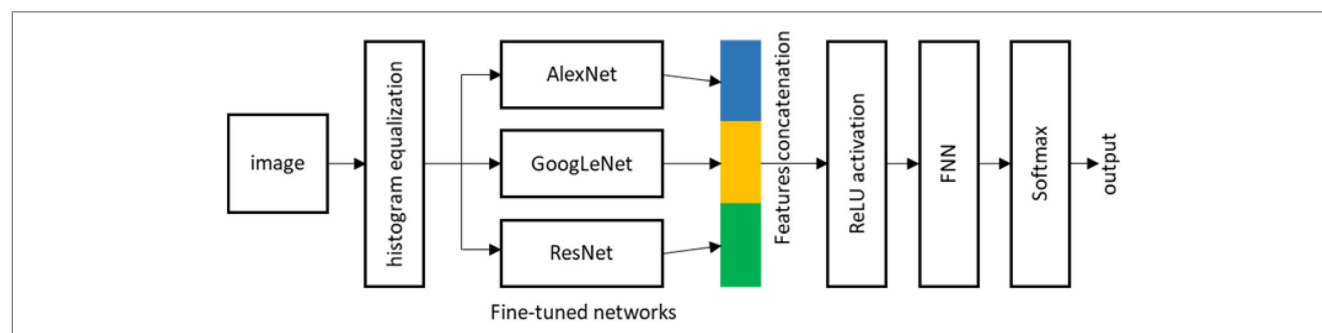
© 2019 Institute of Food Technologists®

Figure 4–CNN-based ensemble network architecture (Pandey et al., 2017). The green, yellow, and blue blocks represent the features extracted by AlexNet, GoogLeNet, and ResNet, respectively. The features for three submodels were concatenated to generate the global features of the whole ensemble network.

78.5% and 94.1% for Top-1% and Top-5% using a fine-tuned deep 50-layer ResNet. It was obvious that the classification results of CNN models on Food-101 dataset were much better than those of traditional methods. UECFood-256 and UECFood-100 are another two publicly available food image databases. A series of experiments were deployed on these two datasets, using similar approaches as those used on Food-101. The performances of these deep learning methods on Food-101, UECFood-256, and UECFood-100 were described in detail in Table 1.

Ciocca and his coauthors contributed two large datasets, Food-527 (Ciocca, Napoletano, & Schettini, 2017) and Food-475 (Ciocca et al., 2018), and introduced ResNet-50 model into food image classification, which achieved the best performance on Food-527, Food-475, Food-50, and VIREO. Deep residual networks were once considered as the best structure for common image classification. Besides, they investigated the use of CNN-based features for food recognition. They compared the effects of different models, different training methods, and different training datasets on food classification accuracy. It was found that the features learned on the largest existing food image database named Food-475 were more representative than features got from other datasets. Thus, they concluded that the dataset containing more features for food image could further promote the food recognition algorithms.

Martinel et al. (2018) were committed to improve the accuracy of food recognition from another point of view. They found that almost all published deep learning methods for food recognition just exploited off-the-shelf deep models or modified the ones without considering the specific characteristics in food images. They first designed a slice convolution unit for extracting common vertical characteristics of food and then added deep residual blocks to make a combination to calculate the classification score. A new CNN structure called WISeR specially used for food recognition was published. Evaluation result demonstrated that the solution achieved the highest Top-1% and Top-5% on Food-101 (90.27%, 98.71%), UECFood-256 (83.15%, 95.45%), and UECFood-100 (89.58%, 99.23%).

Among all the surveyed papers, different datasets, methods, models, and evaluation indicators were used, which bring some difficulties to our summarizing work. Table 1 only concludes the deep learning-based researches related to three popular databases. Moreover, Appendix B summarizes only some popular publicly available food image databases and the best classification achieved on these databases currently.

There were also some papers that designed their experiments on other unpublicized database collected by themselves. Heravi

et al. (2015) designed a network modified from AlexNet and achieved 95% for classification accuracy on a dataset including 1316 images in 13 food categories. Mezgec and Seljak (2017) developed a system named NutriNet, aiming at food and drink identification. The used CNN model was a modification of the AlexNet architecture, which was trained on a training dataset with 225953 images of food and drink items and tested on a detection dataset containing 130517 images. Classification accuracy reached 86.72% and 94.47% for training set and testing set, respectively. Fu et al. (2017) created the ChinFood1000 database and achieved the top-1 and top-5 accuracy of 44.10% and 68.40%, respectively, using ResNet architecture. Herruzo, Bolaños, and Radeva (2016) evaluated the ability of CNN-based classifier for Catalan food. A new dataset called FoodCAT was presented based on Catalan food. Another database used was Food-101. Experiment showed that GoogLeNet that was tuned on Food-101 and FoodCAT and treated by Super-Resolution method achieved the best results for dishes identification (Top-1% of 68.07% and Top-5 of 89.53%) and food categories recognition (Top-1% of 72.29% and Top-5% of 97.07%).

Pandey et al. (2017) developed a multilayered CNN to recognize food. Two different image databases were used, including Food-101 and an Indian food database, between which the later had 50 categories with 100 images of each. The proposed algorithm used AlexNet architecture for deep CNN as baseline and developed a multilayered CNN pipeline to combine the feature outputs of three different subnetworks (AlexNet, GoogLeNet, and ResNet), as shown in Figure 4. Excellent prediction results were obtained with Top-1% of 72.12%, Top-5% of 91.61%, and Top-10 accuracy of 95.95% for Food-101 database and 73.50%, 94.40%, and 97.60% for Indian food database, respectively. The proposed ensemble net outperforms CNN model using one single subnetwork for all the ranks in both the two databases.

Heravi, Aghdam, and Puig (2017) focused on how to create a simple network with fewer parameters while ensuring the performance of the model for the consideration of cost, computing speed, and hardware requirements in practical APP. Usually, network training is carried out by minimizing the error between the output classification score and the ground truth. But they provided a new idea that transfers the knowledge from a large fitted CNN (compressed GoogLeNet architecture) as a trainer to a simple model (with much fewer parameters that could work faster than the trainer CNN) as the trainee on the premise of ensuring the accuracy of the model. The knowledge transferring task was to realize precise approximation of trainee CNN for trainer CNN, see Figure 5 for details. The trainee CNN was expected to
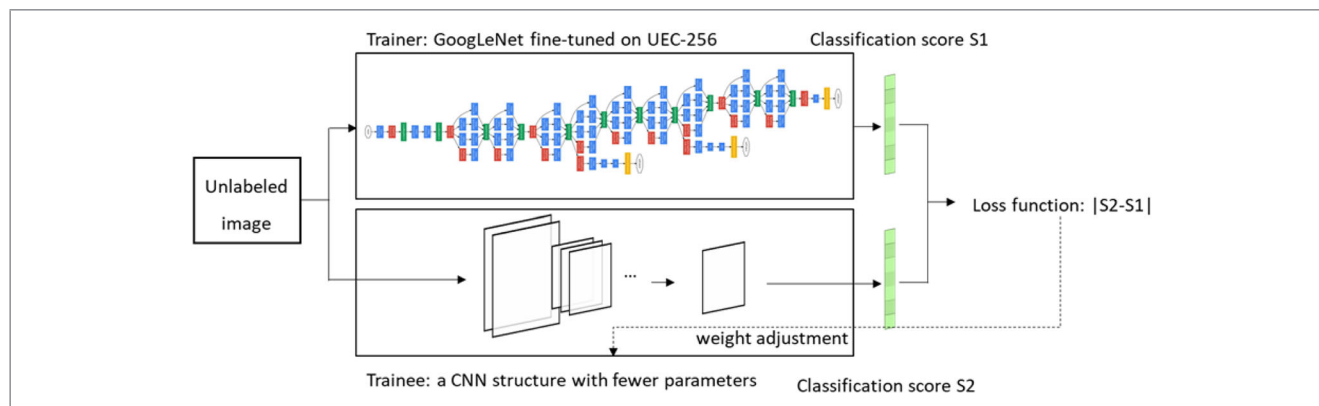
Figure 5–Diagram of knowledge transferring method. The trainer is a trained GoogLeNet and the trainee is a custom CNN with much fewer parameters. S1 is the classification result of the trainer, which can be considered as the ground truth. S2 is the classification result of the trainee. The knowledge transferring task was carried out by minimizing the absolute value of difference between S1 and S2.

provide the same classification scores as the trainer CNN. Thus, it was a function approximation problem rather than classification problem. The trainee was trained with unlabeled nonfood images to approximate the trainer and then further be fine-tuned with labeled food database for food classification. The proposed method achieved 62% for Top-1% on UECFood-256. Although the performance was not very good, this method showed that we could use knowledge transferring method to train a simple network with a lower memory consumption.

In summary, the overall process of learning methods proposed for image-based food recognition in surveyed papers was basically the same. The first step is dataset preparation. Due to the requirement of deep learning for massive data, food images from the Internet or food database with open access are always the first choice for training the model. Next, image preprocessing like normalization, resizing, is followed to weaken the interference caused by nonuniform illumination, resolution inconsistency, and so on. If the dataset is not large enough, data augmentation should be performed to enlarge the dataset by random clipping, rotation, and flipping, to simulate shooting from different perspectives. Generally, experiments based on large open datasets do not consider data augmentation. Then the prepared dataset is always divided into training (or calibration) set for training the network, validation set for fitting the hyperparameters, and evaluation (or testing) set for confirmation of the predictive ability of the model. After the dataset is ready, training task can be carried out as the second step. Existing well-known networks for image classification, such as AlexNet, GoogLeNet, ResNet, and so on, are introduced into food recognition tasks. The majority of papers employed pretrained CNN directly and fine-tuned the model on their database for classification, some researchers made a modification on these networks as the solution, and there were also some authors created new architectures or presented a new method for training procedure specifically targeting at food image recognition. Among the researches discussed in this section, the methods to modify the model structure mainly focused on the combination mode of features. The combination of features of different dataset, extracted from different architectures, was generated as the final feature map for classification. Among the reviewed papers, Top-1% and Top-5% were the most used indicator for performance evaluation. Though excellent results were achieved by summarized solutions, most of the published papers in our survey only tested their solution on the same database (for example, 75% for training and 25% for testing). The generalization ability of the trained

CNN-based model should be examined on different datasets. Furthermore, more information of food like smells and weights can be considered as supplementary to obtain further improvement of recognition accuracy in the future.

Food Calorie Estimation

With the improvement of living standards, dietary health has attracted more and more attention. Many people become interested in keeping track of daily diet to help them control nutrition intake, lose weight, manage their diabetes or food allergies, and improve dietary habits to keep healthy. Food calorie is one of the most concerned indexes. Many mobile APPs have been designed for recording everyday meals including not only food names but also food calorie (Ege & Yanai, 2018; Myers et al., 2015).

Myers et al. (2015) designed a mobile APP named Im2Calories for food calorie estimation from images. The operation process of system could be divided into five steps. The CNN models and training processes of each part are described in Table 2 in detail. First, a fine-tuned GoogLeNet CNN model was utilized to identify whether the image captured was food or not. Second, a fine-tuned GoogLeNet architecture was employed to recognize the food in the picture. Semantic image segmentation was set as the third part of the process to find the location of the food. The next step was physical size prediction for the segmented food. The content of calorie could be calculated according to the list of detected food items, their volume, and calorie density, respectively, as the last step. After the image being parsed into a list of K food items (probability of appearing in the image higher than a certain threshold $\emptyset = 0.5$, K is defined as the number of items satisfying conditions, k is the index), the calculation of total calories was defined as:

$$C = \sum_{k=1}^{K} p(\gamma_k = 1 \,|\, x) C_k \qquad (1)$$

where $p(\gamma_k = 1|x)$ calculates the probability that the kth item appears in the picture x, and C_k denotes the calorie content of menu item k. The presented method achieved good effect of mean absolute error of 152.95 ± 15.61.

Another image-based food calorie estimation research was carried out by Ege and Yanai (2018). A new network called multitask CNN (Figure 6) was adopted for estimating food calorie from a food image. Different from the method used in Myers et al. (2015) which identified food ingredients, volume, etc. step-by-step and

Table 2–The CNN models and dataset used in Im2Calories.

Tasks	Network	Dataset
Food/Nonfood classification	Fine-tuned GoogLeNet (Szegedy et al., 2015)	Food101-Background dataset
Food category recognition	Fine-tuned GoogLeNet	Food101&Food201-MultiLabel[a]
Semantic segmentation	DeepLab (L.C. Chen, Papandreou, Kokkinos, Murphy, & Yuille, 2018)	Food201-Segmented dataset
Volume estimation	Multiscale (Eigen, & Fergus, 2015)	NYU v2 RGBD & GFood3d & NFood-3d

[a] Food201 database is available at https://storage.googleapis.com/food201/food201.zip.
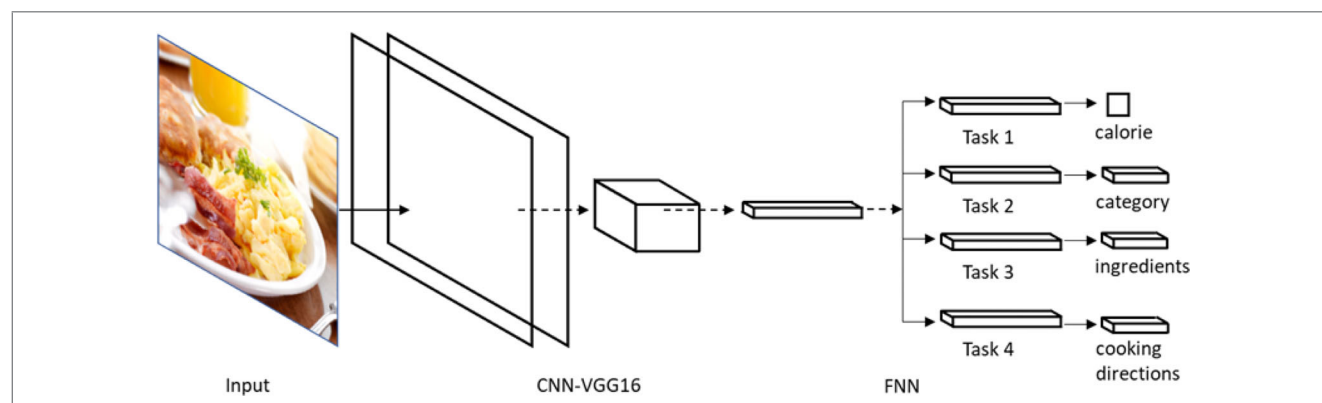

Figure 6–Summary of multitask CNN in (Ege & Yanai, 2018). The multitask CNN used a VGG16 architecture for feature mining and the learned features were fed into four parallel subnetworks to predict the calorie and other attribute of food.

calculated calories finally, the 16-layer VGG network (VGG16)-based multitask CNN predicted food calorie by simultaneously learning food calories, categories, ingredients, and cooking directions on the basis of the existence of correlation between the information mentioned above. Hence, a specific new loss function was defined as

$$L = L_{cal} + \lambda_{cat} L_{cat} + \lambda_{ing} L_{ing} + \lambda_{dir} L_{dir} \qquad (2)$$

where L_{cal} (liner combination of absolute error and relative error), L_{cat}, L_{ing}, and L_{dir} are the loss functions of four separated tasks, and λ_{cal}, λ_{cat}, λ_{ing}, and λ_{dir} represent the weight of each attribute. The training task was carried out by a Japanese dataset and an American dataset collected by authors. The result of multitask CNN achieved 27.4% for related error, 91.2 kcal for absolute and 0.817 for correlation, which was better than independent single-task CNN and image-search–based calorie estimation method.

Food calorie estimation is a more challenging task than food classification. Color and texture information contained in food images is far from enough to estimate calorie content, because the weight or volume of food, cooking directions, and ingredients directly affect the calorie content of food. It is not easy to build a large dataset containing food images, ingredients, cooking methods, and weights (or volumes) labeled with calorie content, which restricts the use of deep learning technology to achieve calorie estimation. Another problem is that the large DNN is difficult to be deployed on a mobile phone to achieve offline detection. In Myers et al. (2015), only the image classification part was embedded in APP, not including image segmentation and volume estimation part. Although such a system can roughly estimate the calorie content of food, it is expected to develop more advanced food information systems that involve enough parameters of a dish to achieve accurate estimation of calorie content of food.

Quality Detection of Vegetables

Vegetables intake is an important part of a healthy diet because of abundant essential nutrients for human beings. During production,

transportation, storage, sales, and other procedures, vegetables are vulnerable to pests, diseases, mechanical damage, and other effects, which reduce their economic value and even affect the health of consumers.

SAE method was adopted by Z. Liu et al. (2018). They developed a classification approach using the stacked sparse autoencoder combined with CNN for cucumbers defect detection based on hyperspectral imaging. The diversity of cucumber surface defects in size and color brings great trouble to the identification method based on average spectrum of the whole sample. Therefore, a CNN model was first used to screen out the defected regions based on the image in RGB channels, by means of using a search window traversing the whole image. If the region inside the search window was judged as defected, the mean spectra of defective region were fed into stacked sparse autoencoder (SSAE) for deep feature representation and classification. Conversely, the mean spectra of whole spectra image were used. The presented CNN–SSAE model outperformed the single SSAE system with 91.1% for accuracy that chose the mean spectra of the whole defected cucumber as the input.

To our knowledge, there are only a few of published papers involving spectral sensing coupled with deep learning for safety and quality evaluation of vegetables, although there are many published researches related to estimation of vegetable yield and detection of vegetable fruit using RGB image sensing and deep CNN. A possible reason is that large datasets of spectra images are difficult to establish because of the low sampling rate and complex pretreatment process, and so on.

Quality Detection of Fruits

Similar to vegetables, fruit is another kind of important food for human being. The production and sales of fruit suffer the same troubles as those of vegetables, such as pests, disease, bruise, and so on. Moreover, fruit is a kind of agricultural product with high value. The freshness, nutrient content, and safety guarantee of fruits are also the issues of concern. Quality detection of fruits and vegetables is currently hot and challenging research areas.

In recent years, deep learning coupled with image processing or spectral sensing method has been widely used as an efficient and nondestructive fruit quality detection tool, to solve the problems such as varieties classification, nutrient content prediction, disease, or damage detection.

Rodriguez, Garcia, Pardo, Chavez, and Luque-Baena (2018) focused on the discrimination of plum varieties (Black Splendor, OwentT, and Angelino) at early maturity stages using deep learning technology. Pictures of samples with different varieties and maturity were captured to build the dataset. Based on the captured images, the proposed method segmented the images first to remove the unwanted background and then the classification was performed by using CNN. AlexNet architecture was chosen as CNN model. The classification accuracy obtained ranges from 91% to 97% in different collected dataset.

Azizah, Umayah, Riyadi, Damarjati, and Utama (2017) applied CNN to detect mangosteen with defected surface. One-hundred twenty RGB images with manual labels were obtained and then cropped and resized to 512 × 512 pixels as the dataset for modeling and evaluation. CNN coupled with fourfold cross-validation was used to solve the problem of binary classification. The proposed method could reach 97.5% classification accuracy for mangosteen defect surface detection.

Tan et al. (2016) aimed at realizing artificial intelligence (AI)-based alerting system for pests and diseases of apple. CNN was applied for recognition of apple skin lesion image collected via an infrared video sensor network. Image preprocessing method employed in this research adjusted the intensity value of each raw image into a certain brightness interval and processed these adjusted images with rotation translation of four different angles, in order to take orientation disturbance and brightness variation into consideration. The size of the image database was expanded to 4000. After dimension reduction by PCA method and being resized to 28*28 pixels, images were trained using a five-layer CNN model defined by authors, and the recognition accuracy was up to 97.5%. The proposed network presented overwhelming odds in parallel with traditional neural networks such as multilayer perceptron (MLP) (68.75%) and KNN (62.50%).

Some damages or lesions on surface of fruit are visible to the naked eye, and features can also be clearly reflected in RGB images. Such detection tasks can be solved by the combination of machine learning and computer vision. However, it is challenging for researchers to precisely detect the mechanical damage under the skin of berries (Blueberry, for example) that composed of deep dark pigments by utilizing RGB imaging technique (Z.D. Wang, Hu, & Zhai, 2018b). Spectroscopy technologies have been frequently applied as nondestructive measurement techniques for representing the internal state of fruit. T. Wang, Chen, Fan, Qiu, and He (2018a) introduced hyperspectral image sensing and deep learning method as the solution to identify the damaged blueberry. Perfect and damaged samples were scanned to capture hyperspectral cube by a hyperspectral transmittance imaging system (wavelengths: 328.82 to 1113.54 nm). Each sample was converted to a size of 32 × 32 with 151 channels via clipping, morphological segmentation, resizing, and subsampling to get the whole database. Adjusted architectures of ResNet and ResNeXt with the fine-tuned parameters were used for classification and achieved 0.8844 and 0.8952, respectively, for accuracy as well as 0.8784 and 0.8905, respectively, for F1-score, which gave a higher precision than conventional classifier sequential minimal optimization (SMO) (0.8082/0.8268), linear regression (LR) (0.7606/0.7796), random forest (RF) (0.7314/0.7529), Bagging (0.7113/0.7339),

and MLP (0.7827/0.7971). The authors pointed out in particular that the size of raw database could not meet the requirement of deep learning, although it was enough for conventional machine learning. In addition to sufficiently large quantity of samples, deep learning methods also require enough feature information of each sample.

Mithun et al. (2018) investigated the ability of CNN method for screening out the artificially ripened banana by hyperspectral sensing and RGB imaging. It should be pointed out that only RGB images (120 images from each class: artificially/normally ripened, for training and 30 images for testing) were used to train and evaluate a modification of AlexNet. The employed model yielded a significant classification accuracy of 90% for this binary classification problem.

The constraints of deep learning-based APP were pointed out by Sun, Wei, Liu, Pan, and Tu (2018). They aimed at developing a system for detection of diseased peaches using spectral analysis. Both partial least-squares discriminant analysis (PLS-DA) and DBN methods showed excellent performance when training based on the data in high-dimensional form (full spectral range of 420 channels, 54 image features, and the fusion of 474 features). However, training on the data after dimension reduction, which consisted of only six optimal features, the PLS-DA model performed better than the deep belief network (DBN) model. Considering the size of storage space occupied by data, computing speed, and other requirements for hardware, simple models and spectra of less channels have more advantages for industrial APP.

A new deep learning architecture was established to estimate the firmness and soluble solid content of pears by Yu, Lu, and Wu (2018). One-hundred eighty pears were place in chambers, 15 of which were taken out every other day to obtain VIS/NIR spectral data, and labeled with values of their reference firmness and soluble solid content. SAE was trained to extract high-level features of raw spectra. Then, the extracted features were put into an FNN for estimation of these two attributes of pears. When training the SAE model, pixel-level spectra in region of interest (ROI) were used to ensure that the dataset was big enough. The mean spectra were calculated and considered as the input of the trained SAE-FNN model for reducing the impact of interference. The results obtained with R_P^2 of 0.890, RMSEP of 1.81N, RPD$_P$ of 3.05 for firmness, and R_P^2 of 0.921, RMSEP of 0.22%, and RPD$_P$ of 3.68 for soluble solid content indicated that the combination of deep learning and visible and near infrared (VIS/NIR) spectral sensing can be used as an effective nondestructive method for quality detection of fruit.

Moreover, W.S. Zhang et al. (2018a) integrated AI technology into the design of smart refrigerator to achieve fruit recognition. The proposed approach focused on distinguishing different kinds of fruits and identifying different individuals in the image. The dataset including pictures of fruit captured in a refrigerator and photo resources from the Internet was used to train SSD models (ResNet, VGG16, and 19-layer VGG network [VGG19]). The output of these submodels would be concatenated and further fed into a back propagation (BP) neural network. Furthermore, weight information of a fruit was collected to help the recognition process. The combination of multiple source information and the fusion of multiple CNN architecture showed better recognition accuracy than use three models separately without weight information.

These surveyed studies indicated that the types of fruits and some physical and chemical indicators of them (such as firmness, nutrient content, damage degree, disease degree, natural maturity, and so on) will be reflected in the RGB image or spectral
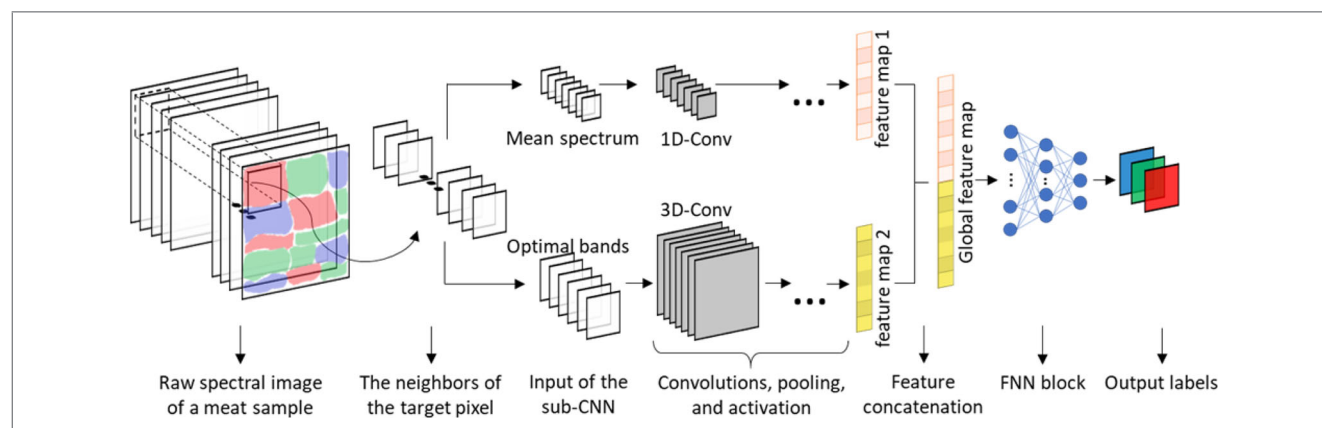
Figure 7–A deep CNN structure for spectral–spatial features extraction. The input of the network is the neighbors of the target pixel, a small region of a spectral image, which were fed into two parallel subnetworks. The target region of a spectral image was used to calculate the mean spectrum and treated by one-dimensional CNN in the first subnet, while be reduced the number of channels to compressed input size and processed by three-dimensional CNN in another subnet. The abstracted features from the subnets were concatenated together and classified by an FNN block. After all the pixels are traversed, the prediction result of a spectral image would be obtained.

information of samples, which can be learned and discriminated by deep learning models to predict the parameters related to quality and safety. In Appendix A, we summarized the technical details (including the target problem, composition of dataset, pretreatment methods, models and frameworks, and also the performance comparison with other algorithms) of the researches mentioned in this section and the previous section. Deep learning methods showed better performance than traditional data analysis methods and deserve further study in the future for quality detection of fruits.

Quality Detection of Meat and Aquatic Products

Aquatic products (such as fish, shrimp, and so on) and meat (such as pork, mutton, beef, and so on) become an important part of human diet for protein supply. Measurement of multiple chemical indicators is conducive to food safety supervision in aquatic production process. In the past few years, spectral sensing and machine intelligence have been widely used for rapid nondestructive quality detection of aquatic products (for example, chemical properties of fish muscle prediction) (J.H. Cheng & Sun, 2017). Data analysis algorithms such as partial least squares regression (PLSR), SVM, LR, and so on, act as powerful tools for quality classification, freshness, and nutrient prediction based on spectral data of samples. Recently, with sufficient evidence of strong feature learning and data analysis ability, deep learning methods begin to be introduced into this area to replace traditional machine learning methods.

Yu, Tang, Wu, and Lu (2018) investigated deep learning model to deal with visible/near-infrared hyperspectral data of shrimps for freshness prediction of shrimp. SAE model was chosen to obtain the deep features of the samples, and logistic regression was utilized to classify the freshness grade of shrimp using the obtained the deep features. The proposed method provided a good result for shrimp freshness grade classification (96.55% and 93.97% in calibration and prediction sets, respectively).

Yu, Wang, Wen, Yang, and Zhang (2019) also studied how to correlate the hyperspectral data for determining total volatile basic nitrogen (TVB-N) content in shrimp using deep learning. For dataset preparing, a batch of live shrimps was treated in different ways to obtain samples with different TVB-N concentration gradients, which were scanned by a spectrometer. The TVB-N

content of each sample was obtained by chemometrics method. In modeling procedure, successive projections algorithm and SAE algorithms were used as feature extractors for comparison. Moreover, three kinds of regression algorithms including least-squares support vector machine (LS-SVM), PLSR, and multiple linear regression (MLR) were used for TVB-N content prediction. It was found that the combination of SAEs and LS-SVM model showed the best performance ($R_P^2 = 0.921$, RMSEP = 6.22 mg/100g, RPD = 3.58) with a low time consumption (3.9 ms).

Concerning meat quality analysis, Al-Sarayreh et al. (2018) provided a new architecture for adulteration identification in red-meat products using hyperspectral imaging. Different kinds of meat was prepared, pretreated, and subsequently put into frame-style containers. Hyperspectral images of meat samples, a mixture of lamb, pork, and fat, were scanned to build up the dataset. As shown in Figure 7, the structure of proposed deep CNN contained two subnets, which allowed two types of data input including spectral data (mean spectrum of the target region) and spatial information (a combination of adjacent regions of the target pixels in six optimal bands). Spectral data were fed into the first sub-CNN based on one-dimensional convolutions to obtain the spectral features, and spatial information was processed by another three-dimensional convolution based sub-CNN. Feature maps extracted from sub-CNNs were concatenated to form the global feature map, which acted as the input of the followed FNN block to complete the prediction. Furthermore, L2 regularization method was adapted in the proposed model to avoid overfitting problem. Considering the texture information and spectral features of hyperspectral images synthetically, the performance of the presented method outperformed the traditional machine learning algorithm assisted by handcrafted features with an average overall accuracy of 94.4%.

In the aspect of fish and meat quality analysis, deep learning technology coupled with spectroscopy technology could excellently accomplish food quality detection according to the detected internal and external characteristics of food. There are few food-related cases using deep learning to process the spectral information from close range scanning, whereas the method of deep learning for hyperspectral data processing has been widely used in remote sensing field, mainly in spectral image segmentation (Kemker, Luu, & Kanan, 2018), spectral image denoising (Yuan, Zhang, Li, Shen, & Zhang, 2019), spectra unmixing (X.R. Zhang, Sun,

Zhang, Wu, & Jiao, 2018b), target detection (H. Chen, Zhang, Ma, & Zhang, 2019), and so on. The successful cases of deep learning for hyperspectral image processing in remote sensing indicate that deep learning methods have the potential to play a great role in food industry. It is worth studying how to introduce these advanced methods used for remote sensing hyperspectral image processing into food domain, such as nutrient proportion estimation, deterioration region recognition, poor food detection, image segmentation of different components in meat, and aquatic products in future research.

Food Supply Chain

The food supply chain is a complex system consisting of multiple economic stakeholders from primary producers to consumers (including farmers, production factories, distributors, retailers, and consumers) (Mao, Wang, Hao, & Li, 2018). It is hard for regulators, such as governments, to obtain reliable food information due to the unreliable information from the supply chain, which can easily lead to food fraud and food safety problems. Mao et al. (2018) presented a credit evaluation system based on blockchain for food supply chain using a deep learning network named LSTM. The evaluation task was carried out by analyzation of credit evaluation texts. Text data like "The fruit does not look very fresh" were labeled as "negative," and the sentence such as "the quality is good" has a "positive" label. Sentence feature extraction was performed by LSTM and these features act as the input of DNN-based classifier. The proposed method showed approximately 90% classification accuracy on a Chinese text dataset, which was beyond the reach of traditional methods such as SVM and navie Bayes.

The research in (Mao et al., 2018) solved a problem that how to transform large number of credit evaluation text data into some simple evaluation indicators. Similarly, another article (Kim, 2014) introduced a sentence classification method using deep learning, which can be tested using the dataset mentioned in Mao et al. (2018) for comparison. Such research relies heavily on big data, so much work (such as dataset in the type of audio, text, and so on, for food domain should be collected, for example) remains to be done in the future.

Food Contamination

It is possible for food to be contaminated by poisonous and harmful substances due to the effect of environment or human factors on food in any process of planting or feeding, growth, harvesting or slaughtering, processing, storage, transportation and sale, and so on, before consumption. Food contamination can lead to gastrointestinal infectious diseases and is harm to human health, and thus attracts increasing attention all over the world.

Song, Zheng, Xue, Sheng, and Zhao (2017) researched an evolution method for predicting morbidity of gastrointestinal infections by food contamination using DNN. The research was designed for the morbidity prediction of gastrointestinal infectious diseases using a large number of contaminant-related information (from 227 types of contaminants in different concentrations and 119 types of food widely consumed in the investigated region) acquired in the current week as well as the previously recorded morbidity information. The big data of contaminant indexes in the dataset were provided by food supervision departments of the target region in central China, and the morbidity data ware supported by departments of gastroenterology of the hospitals in the corresponding researched region. Deep denoising autoencoder, which is a structure similar to SAE with multiple hidden layers, was constructed to extract hidden features for contaminant indexes and

the extracted representation was used for the supervised learning to predict the morbidity. The presented ecogeography-based optimization (EBO)-based method for the calibration of DDAE model achieved better performance than conventional ANN and deep denoising autoencoder trained by other algorithms, with mean average percentage error of 21.16% and success rate of 58.50%. It was concluded by authors that the deep learning model demonstrated strong capabilities to deal with some incomplete and imperfect information. In the future, more potential contamination features need to be taken into consideration.

There have been many published methods for detection of food contamination utilizing traditional machine learning algorithms (Bisgin et al., 2018; Ropodi, Panagou, & Nychas, 2016; Ravikanth, Jayas, White, Fields, & Sun, 2017). There are potential possibilities for deep learning to replace the traditional machine learning method to achieve better detection results for food contamination in different food production procedures.

Application Potential of Deep Learning in Chemometrics and Sensometrics Toward Food

In the case of food quality guaranteed, people are willing to pursue better sensory experience in food consumption. Thus, the prediction of sensory characteristics of food (Caballero, Antequera, Caro, Duran, & Perez-Palacios, 2016) as well as eating environment and context in sensory and consumer researches (Stelick, & Dando, 2018) begin to attract more research efforts.

Chemometrics (a subject aiming to establish the relationship between the measured chemical parameters and the state of the object by statistical or mathematical methods) and sensometrics (which can be understood as a subject to link sensory parameters and the internal characteristics of the object via similar approaches in chemometrics studies) toward food have been studied to provide approaches and tools to the researches related to production, quality assurance, and consumer choice, which are crucial to food sensory and consumer researches. Some researchers tried to find the relationship between food sensory parameters and catering hedonic data using traditional data processing methods such as PLSR, SVM, neural networks, and so on. The mentioned algorithms were employed as the data mining tools to build up the linear or nonlinear model to estimate hedonic data using food sensory parameters (Luaces, Díez, Joachims, & Bahamonde, 2015). The sensory characteristics of food itself have a great impact on dietary consumption. Some advanced sensing technologies have been used to predict food sensory data and physical and chemical parameters in a nondestructive way (Caballero et al., 2016, 2017; Pérez-Palacios et al., 2017). Traditional machine learning methods including MLR and isotonic regression methods were deployed to establish the mapping relationship between texture features and sensory characteristics, as well as physicochemical parameters of food.

It is illustrated in Qannari (2017) that sensometrics is the intersection of psychometrics, biometrics, chemometrics, and so on, but lags behind the latter disciplines. In addition to following development of statistical methods, researchers should be keenly aware of advances in other disciplines especially big data analysis and deep learning. To our knowledge, sensory and consumer researches related to food are still stuck in the traditional data analysis methods, including image feature extraction, statistical analysis, and so on, while deep learning methods for data/feature mining were hardly ever used.

It is worth mentioning that strong commonality can be found in different APP fields related to food science such as preference mapping and sensory parameters estimation mentioned in this chapter, calorie estimation introduced in chapter "Food calorie estimation," and physicochemical parameters prediction for fruits described in chapter "Quality detection of fruits." These studies were carried out based on the combination of data analysis algorithms, computer vision, sensometrics, chemometrics, and so on. Among them, data processing methods (mainly machine learning) played the same role in these APPs, being employed for information mining in big data and finding the direct or potential correlation between food sensory parameters, physicochemical parameters, calorie content, preference of consumers, and a series of digital information of food collected by sensing technology and equipment (RGB images, magnetic resonance images, spectral images, smell information, and so on).

According to the experience in recent research in deep learning for food classification, quality detection, and so on, DNNs can be further extended to sensometrics- and chemometrics-related researches toward food with its powerful feature representation ability to achieve automatic data mining from various data sources. We believe that deep learning method has the potential to promote the development of chemometrics, sensometrics, sensory, and consumer sciences toward food, with strong evidence of its powerful ability that outperforms traditional machine learning methods and its success in big data analysis, automatic feature mining, and modeling mention in this survey.

Challenges and Future Perspective of Deep Learning in Food Domain

The most significant advantage of deep learning technology is feature learning. Traditional machine learning approaches use raw data as input or deal with classification tasks based on hand-crafted features. Deep learning methods can learn representational features from the dataset during the training process, and demonstrate stronger ability than traditional methods. Another characteristic of deep learning is its ability of transfer learning. In the researches mentioned in the section "Food recognition and classification," we found that most of them exploited pretrained CNN models based on large dataset (such as ImageNet) and fine-tuned the models on their target datasets, which could reduce the difficulty and time consumption for training a model (even based on a much smaller dataset). Moreover, there were also some authors deploying features extracted by a CNN to train another classifier like SVM, to transfer the knowledge from CNN model to a new classifier.

Different from conventional data analysis methods, deep learning technology involves more complex model structure and computational efforts that have once limited its development and APP. Nowadays, benefiting from the global emphasis on deep learning and the contributions of scientists, many tools have emerged to help researchers get a quick start to make a deep learning–based APP. As for software support, we would like to list several popular frameworks designed for reducing programming difficulty for researchers: Theano (deeplearning.net/software/theano/), Tensorflow (tensorflow.google.cn), Caffe (caffe.berkeleyvision.org), Pytorch (pytorch.org), MXNet (mxnet.incubator.apache.org), Keras (keras.io), and MatConvNet for Matlab (www.vlfeat.org/matconvnet). It seems difficult to program complex neural network models, but with the help of these existing frameworks, we can quickly build the required network model by stacking some duplicate network structures by calling encapsulated function interfaces. As for hardware support, graphics processing unit (GPU) coupled with Compute Unified Device Architecture (CUDA) Toolkit and the NVIDIA CUDA Deep Neural Network library (cuDNN, a GPU-accelerated library of primitives for DNNs) produced by NVIDIA company can provide hardware and software acceleration for deep learning computation. These toolkits accelerate widely used deep learning frameworks mentioned above. Software and hardware acceleration tools greatly shorten the computing time and have the potential to meet the requirements of real-time data processing.

However, the fact that deep learning has shortcomings is undeniable. Due to long training time as well as hardware restrictions, added with the high complexity and numerous hyperparameters of the model, the optimization tasks would be very complicated and time-consuming. The mentioned GPUs for computing acceleration and matched processors and other hardware are very expensive. It will take much longer time to train a DNN only using CPUs as the computational hardware. Furthermore, deep learning requires big data for training, and the acquisition of reliable big dataset is another difficult problem. Data collection and annotation will take a lot of time and energy. Some open datasets for academic research and challenge competition were collected and manually labeled by experts or volunteers, even could be directly downloaded from the Internet by machines; thus, there would be some mistakes inevitably. Moreover, it should be noted that the trained network can only understand the features of the dataset used for training. Some published datasets describe incomplete information for the target problem. As an example, UECFood-256 database is a combination of large quantities of Japanese food images, so the model trained on this database will be difficult to accurately identify food from other countries and regions. To achieve a more accurate and universal food recognition model, larger datasets containing food images from all over the world should be constructed. In section "Deep learning APPs in food," we listed the APPs of deep learning in food domain. It is not difficult to find that food identification, fruit, vegetable, fish, and meat quality detection tasks have begun to use deep learning technology, but only a very few of articles can be found that applied deep learning for food calorie estimation, food supply chain, and food contamination problems, and also for menu recognition (Lee, Chiu, & Chang, 2017), nutrition measurement (Pfisterer, Amelard, Chung, & Wong, 2018), food safety risk early warning (Geng, Shang, Han, & Zhong, 2019), personalized diet recommendation (C.H. Chen, Karvela, Sohbati, Shinawatra, & Toumazou, 2018), dining experience (Naritomi, Tanno, Ege, & Yanai, 2018), visual modeling of food (P.Y. Chen et al., 2019), and so on, which we did not describe in detail in the survey.

In the case of food recognition problems, although there were dozens of papers reporting their APP of deep learning, RGB image information was the only basis used to distinguish food types. In a portion of the aforementioned studies for food classification, the contributors were mainly scholars of computer science and image processing; thus, more attention was attracted by the general features of images instead of the specific features of food images. As for food quality detection, in addition to images, characterization methods such as hyperspectral imaging and thermal imaging have been used to reflect the intrinsic information of food. Due to the large size of original spectral images and the large storage occupation, the mean spectrum of the ROI is usually used to represent this region, to fit the computing ability and storage capacity of the hardware. On the basis of the mean spectrum of a sample, by choosing an apple as an example, deep learning method could only
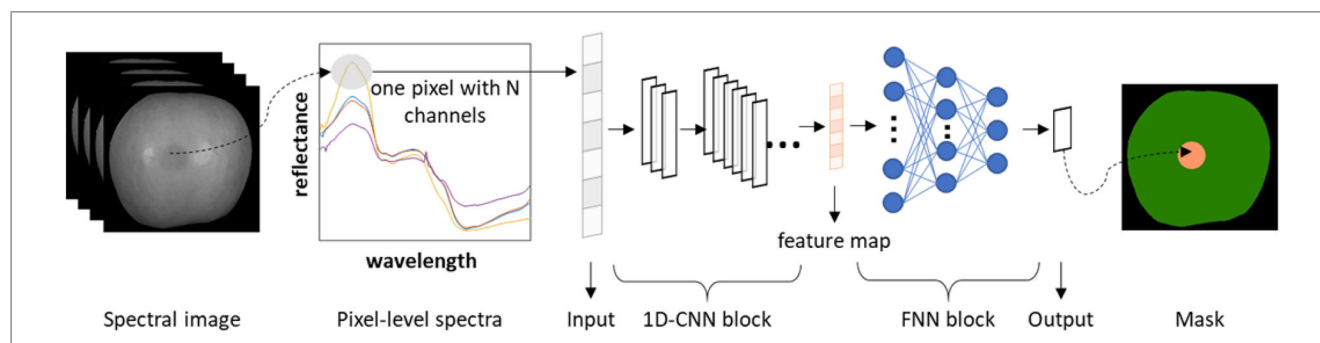
Figure 8–A CNN architecture for pixel-level classification. Here, we take a spectral image of a bruised apple as an example. Each pixel in the spectral image would be put into the CNN model and processed by one-dimensional convolutions to get the features. Then, the binary classification would be performed by using an FNN block. The classification results of each pixel could be put together to generate a mask. The regions with "damaged" label could be considered as the bruised areas.

predict whether the target region is defected/sick/insect damaged or not, and it was hard to judge where the particular location is. It can be considered to extract some key information to reduce the size of the spectral images. One way is to find some optimal bands that can best reflect the differences among samples, and recombine the corresponding layers as new images, which mainly considers spatial information. Another idea is to train the model based on the pixel-level spectra, and finally reconstruct the prediction label of each pixel into a mask as output, as seen in Figure 8. The implementation of one-dimensional convolution can be found in Qiu et al. (2018) and N. Wu et al. (2018). If the problem to be solved has little relevance to spatial or texture information, this method can be a good choice to calculate the predicted values of each point separately to overcome the limitation of hardware storage. Moreover, the combination of spatial and spectral features can be utilized to solve specific problems, similar to the solution in Al-Sarayreh et al. (2018).

More types of information of food are expected to be used for training a deep learning model. As human beings, we will judge what kind of food it is and evaluate its quality according to not only the appearance of food but also its weight, smell, touch, firmness, taste, sounds after being knocked, and so on. Currently, there are many kinds of sensors available for measurement of these information, such as electronic balance, electronic nose, vibration sensor, sound sensor, and so on, for nondestructive measurement, and there also are a series of advanced detection technologies such as hyperspectral imaging, Terahertz, fluorescence spectroscopy, thermal imaging technology, and so on, to obtain interior information of food. Multisource data fusion have not been fully utilized in food quality and safety evaluation with deep learning. One example was given in this paper that considered the combination of image and weight information of fruits to give a more accurate identification of fruits (Zhang et al., 2018a). Multisource data (or features) fusion based on more types of data from advanced sensors could be further used to achieve a more comprehensive and accurate evaluation of the food. In the future researches, the food to be analyzed is not only limited to aquatic products and meat, vegetables, and fruits, but also expended to liquid foods, such as milk, beverages, and so on.

In addition, time dimension should be taken into account. Some problems are difficult to describe with static data. Taking the evaluation of dough fermentation as an example, images or other data representing the current fermentation state are not enough to illustrate the problem. The parameters of the whole fermentation process at different times must be used. Other examples, such as

food recommendation, menu recognition, food texture detection based on vibration information, and so on, rely on time-varying data or sequential data. RNN and LSTM models coupled with classifier or regression algorithms are expected to be utilized in the future work in food domain.

Moreover, we believe that handheld smart devices and mobile APPs for food identification, safety, and quality evaluation will appear in the near future. Due to the improvement of hardware computing ability, the training task can be completed in a much shorter time using NVIDIA GPUs (www.nvidia.com) assembled on PC or remote servers. According to the surveyed articles, some authors had further integrated food recognition methods into mobile apps after finishing the modeling task. Two problems should be considered: First, how to embed the model into mobile app. Due to the massive model structure and large number of parameters, mobile phones could not provide enough memory space and computing capability for the APP. In the article Myers et al. (2015), only food recognition function was implemented in the APP, excluding image segmentation and calorie estimation. Zhang et al. (2018a) considered to assign computing tasks to a cloud server. The captured picture would be transmitted to the remote server for analysis and the result would return to the APP. Although training task would take a lot of time even using GPUs, the predicting task could be finished quickly, which was enough for designing APPs. The second problem is how to realize miniaturization of sensing equipment. Except the large spectrometers used in the laboratory, some portable sensing devices combined with traditional machine learning algorithms and Internet of things technology have been applied in fruit quality evaluation, egg freshness prediction (Coronel-Reyes, Ramirez-Morales, Fernandez-Blanco, Rivero, & Pazos, 2018), and achieved acceptable results (Wang et al., 2018a). Miniaturized hardware computing platforms are also beginning to be used, such as NVIDIA Jetson TX2 (Partel, Kakarla, & Ampatzidis, 2019), which is expected to be utilized to realize local accelerated computation without the help of the internet and cloud servers.

Specially for food recognition challenges, some free downloadable food image databases for academic use, added with the download address and the state-of-the-art performance achieved on the corresponding database, were listed in Appendix B to attract more challengers and researchers to expand these datasets, and propose more advanced algorithms/networks to improve classification accuracy, and further put into practical APP.

The rapid development of Internet, social media, mobile app, and other technologies has provided more diverse approaches for

data collection that allows more people to participate and contribute food information such as images and text descriptions, to promote the emergence of much larger datasets in the future. Researchers and institutes all over the world conducted food quality and safety inspection by their own datasets. The power of a single person, research team, or institution for data collection is limited. It is expected to integrate food-related databases acquired by modern sensors and instruments from worldwide users, researchers, and institutes into the large global databases. With the advantage of deep learning, these datasets can be analyzed efficiently, which would benefit researchers and institutes of food domain.

Conclusion

In this paper, we investigated a large number of latest articles related to the APP of deep learning in food, described in detail the proposed structure, training methods, and the final evaluation result of DNNs used to process food image, spectrum, text, and other information in each surveyed article. In the aspect of performance, we compared the deep learning with other existing popular methods, and found that the deep learning method achieves better results than other methods in these reviewed studies. We concluded the advantages and disadvantages of deep learning methods and made a detailed discussion of the challenges and future perspective of deep learning in food domain. To authors' knowledge, it is the first survey on the APPs of deep learning in the food domain. The purpose of this review is to encourage researchers and workers in this field to perform more experiments on food with deep learning methods, to present precise solutions for classification or regression problems and put them into practice for the benefits of food quality and safety inspection for human dietary health. At last, we recommend that (1) the combination of deep learning and multisource data fusion including RGB images, spectra, smell, taste, and so on, would be considered to make a more comprehensive assessment of food, (2) the development of full-automatic information acquisition equipment/systems with stable signal output for food and global food data sharing platforms should be studied in the future, since it is still very hard to obtain big data related to food due to the usage of semiautomatic or even manual information acquisition tools and incomplete data management and sharing platforms, (3) the potential of deep learning technology in data mining can be evaluated in food related areas rarely explored such as food sensory and consume, food supply chain, and so on, and (4) successful cases of deep learning such as in food (such as food image recognition, intelligent recipe recommendation APP, and fruit quality evaluation system) can be further transformed into practical products.

Acknowledgments

Author Contributions

L. Zhou, C. Zhang, Y. He, Z. Qiu, and F. Liu conceived and designed the survey. L. Zhou and C. Zhang collected data, interpreted the results, and drafted the manuscript. Y. He, Z. Qiu, and F. Liu critically revised the draft.

Abbreviations

AI	artificial intelligence
ANN	artificial neural networks
APP	application
BoW	bag-of-visual-words
CNN	convolutional neural network
CUDA	Compute Unified Device Architecture
DBN	deep belief network
DNN	deep neural network
EBO	ecogeography-based optimization
EMP-SVM	extended morphological profile-SVM
FCN	fully Convolutional Network
FNN	fully connected network
GPU	graphics processing unit
HOG	histogram of oriented gradient
ICA	independent component correlation algorithm
KNN	k-nearest neighbor
LR	linear regression
LS-SVM	least-squares support vector machine
LSTM	long short-term memory
MLP	multilayer perceptron
MLR	multiple linear regression
NB	navie Bayes
PC	principal component
PCA	principal component analysis
PLS-DA	partial least-squares discriminant analysis
PLSR	partial least squares regression
R-CNN	region-based CNN
ResNet	residual neural network
RF	random forest
RNN	recurrent neural network
ROI	region of interest
SAE	stacked autoencoders
Seq2Seq	sequence to sequence
SIFT	scale-invariant feature transform
SMO	sequential minimal optimization
SPA	successive projections algorithm
SSAE	stacked sparse auto-encoder
SSD	single-shot multibox detection
SURF	speedup robust features
SVM	support vector machine
Top-1%	Top-1 classification accuracy
Top-5%	Top-5 classification accuracy
TVB-N	total volatile basic nitrogen
VGG	visual geometry group network
VGG16	16-layer VGG network
VGG19	19-layer VGG network
WISeR	wide-slice residual networks
WT	wavelet transform

References

Abadi, M., Agarwal, A., Barham, P., Brevdo, E., & Zheng, X. (2016). TensorFlow: Large-scale machine learning on heterogeneous distributed systems. Retrieved from http://arxiv.org/abs/1603.04467

Ahmed, A., & Ozeki, T. (2015). Food image recognition by using Bag-of-SURF features and HOG Features. In *Proceedings of the 3rd International Conference on Human-Agent Interaction* (pp. 179–180). https://doi.org/10.1145/2814940.2814968

Al-Sarayreh, M., Reis, M. M., Yan, W. Q., & Klette, R. (2018). Detection of red-meat adulteration by deep spectral-spatial features in hyperspectral images. *Journal of Imaging*, *4*(5), 20. https://doi.org/10.3390/jimaging4050063

Azizah, L. M., Umayah, S. F., Riyadi, S., Damarjati, C., & Utama, N. A. (2017). Deep learning implementation using convolutional neural network in mangosteen surface defect detection. In *2017 7th IEEE International Conference on Control System, Computing and Engineering (ICCSCE)* (pp. 242–246). https://doi.org/10.1109/ICCSCE.2017.8284412

Barbin, D. F., Felicio, A., Sun, D. W., Nixdorf, S. L., & Hirooka, E. Y. (2014). Application of infrared spectral techniques on quality and compositional attributes of coffee: An overview. *Food Research International*, *61*, 23–32. https://doi.org/10.1016/j.foodres.2014.01.005

Bay, H., Ess, A., Tuytelaars, T., & Van Gool, L. (2008). Speeded-up robust features (SURF). *Computer Vision and Image Understanding*, *110*(3), 346–359. https://doi.org/10.1016/j.cviu.2007.09.014

Bisgin, H., Bera, T., Ding, H. J., Semey, H. G., Wu, L. H., . . . Xu, J. (2018). Comparing SVM and ANN based machine learning methods for species identification of food contaminating beetles. *Scientific Reports*, *8*, 12. https://doi.org/10.1038/s41598-018-24926-7

Bossard, L., Guillaumin, M., & Gool, L. V. (2014). Food-101-Mining discriminative components with random forests. In D. Fleet, T. Pajdla, B. Schiele, & T. Tuytelaars (Eds.), *Computer Vision - ECCV 2014, Pt Vi* (Vol. 8694, pp. 446–461). https://doi.org/10.1007/978-3-319-10599-4_29

Brosnan, T., & Sun, D. W. (2004). Improving quality inspection of food products by computer vision - A review. *Journal of Food Engineering*, *61*(1), 3–16. https://doi.org/10.1016/s0260-8774(03)00183-3

Caballero, D., Antequera, T., Caro, A., Duran, M. L., & Perez-Palacios, T. (2016). Data mining on MRI-computational texture features to predict sensory characteristics in ham. *Food and Bioprocess Technology*, *9*(4), 699–708. https://doi.org/10.1007/s11947-015-1662-1

Caballero, D., Antequera, T., Caro, A., Ávila, M. D. M., Rodríguez, P. G., & Perez-Palacios, T. (2017). Non-destructive analysis of sensory traits of dry-cured loins by MRI-computer vision techniques and data mining. *Journal of the Science of Food and Agriculture*, *97*(9), 2942–2952. https://doi.org/10.1002/jsfa.8132

Chen, C. H., Karvela, M., Sohbati, M., Shinawatra, T., & Toumazou, C. (2018). PERSON-personalized expert recommendation system for optimized nutrition. *IEEE Transactions on Biomedical Circuits and Systems*, *12*(1), 151–160. https://doi.org/10.1109/tbcas.2017.2760504

Chen, H., Zhang, L. B., Ma, J., & Zhang, J. (2019). Target heat-map network: An end-to-end deep network for target detection in remote sensing images. *Neurocomputing*, *331*, 375–387. https://doi.org/10.1016/j.neucom.2018.11.044

Chen, J. J., & Ngo, C. W. (2016). Deep-based ingredient recognition for cooking recipe retrieval. In *Proceedings of the 24th ACM International Conference on Multimedia* (pp. 32–41). https://doi.org/10.1145/2964284.2964315

Chen, L. C., Papandreou, G., Kokkinos, I., Murphy, K., & Yuille, A. L. (2018). DeepLab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected CRFs. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *40*(4), 834–848. https://doi.org/10.1109/tpami.2017.2699184

Chen, P. Y., Blutinger, J. D., Meijers, Y., Zheng, C. X., Grinspun, E., & Lipson, H. (2019). Visual modeling of laser-induced dough browning. *Journal of Food Engineering*, *243*, 9–21. https://doi.org/10.1016/j.jfoodeng.2018.08.022

Cheng, G., & Han, J. W. (2016). A survey on object detection in optical remote sensing images. *ISPRS Journal of Photogrammetry and Remote Sensing*, *117*, 11–28. https://doi.org/10.1016/j.isprsjprs.2016.03.014

Cheng, J. H., & Sun, D. W. (2017). Partial least squares regression (PLSR) applied to NIR and HSI spectral data modeling to predict chemical properties of fish muscle. *Food Engineering Reviews*, *9*(1), 36–49. https://doi.org/10.1007/s12393-016-9147-1

Ciocca, G., Napoletano, P., & Schettini, R. (2017). Food recognition: A new dataset, experiments, and results. *IEEE Journal of Biomedical and Health Informatics*, *21*(3), 588–598. https://doi.org/10.1109/jbhi.2016.2636441

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. https://doi.org/10.1016/j.cviu.2018.09.001

Coronel-Reyes, J., Ramirez-Morales, I., Fernandez-Blanco, E., Rivero, D., & Pazos, A. (2018). Determination of egg storage time at room temperature using a low-cost NIR spectrometer and machine learning techniques. *Computers and Electronics in Agriculture*, *145*, 1–10. https://doi.org/10.1016/j.compag.2017.12.030

de Ridder, D., Kroese, F., Evers, C., Adriaanse, M., & Gillebaart, M. (2017). Healthy diet: Health impact, prevalence, correlates, and interventions. *Psychology & Health*, *32*(8), 907–941. https://doi.org/10.1080/08870446.2017.1316849

Deng, J., Dong, W., Socher, R., Li, L. J., Li, K., & Li, F. F. (2009). ImageNet: A large-scale hierarchical image database. In *2009 IEEE Conference on Computer Vision and Pattern Recognition*, Miami, FL (pp. 248–255). https://doi.org/10.1109/CVPR.2009.5206848

Duchi, J., Hazan, E., & Singer, Y. (2011). Adaptive subgradient methods for online learning and stochastic optimization. *Journal of Machine Learning Research*, *12*, 2121–2159.

Ege, T., & Yanai, K. (2018). Image-based food calorie estimation using recipe information. *IEICE Transactions on Information and Systems, E*, *101D*(5), 1333–1341. https://doi.org/10.1587/transinf.2017MVP0027

Eigen, D., & Fergus, R. (2015). Predicting depth, surface normals and semantic labels with a common multi-scale convolutional architecture. In *2015 IEEE International Conference on Computer Vision* (pp. 2650–2658). https://doi.org/10.1109/ICCV.2015.304

Fu, Z. H., Chen, D., & Li, H. Y. (2017). ChinFood1000: A large benchmark dataset for Chinese food recognition. In D. S. Huang, V. Bevilacqua, P. Premaratne, & P. Gupta (Eds.), *Intelligent Computing Theories and Application, ICIC 2017, Pt I* (Vol. 10361, pp. 273–281). https://doi.org/10.1007/978-3-319-63309-1_25

Geng, Z. Q., Shang, D. R., Han, Y. M., & Zhong, Y. H. (2019). Early warning modeling and analysis based on a deep radial basis function neural network integrating an analytic hierarchy process: A case study for food safety. *Food Control*, *96*, 329–342. https://doi.org/10.1016/j.foodcont.2018.09.027

Giovany, S., Putra, A., Hariawan, A. S., & Wulandhari, L. A. (2017). Machine learning and SIFT approach for Indonesian food image recognition. *Discovery and Innovation of Computer Science Technology in Artificial Intelligence Era*, *116*, 612–620. https://doi.org/10.1016/j.procs.2017.10.020

Glorot, X., & Bengio, Y. (2010). Understanding the difficulty of training deep feedforward neural networks. In *Proceedings of the 13th International Conference on Artificial Intelligence and Statistics* (Vol. 9, pp. 249–256).

Granato, D., Santos, J. S., Escher, G. B., Ferreira, B. L., & Maggio, R. M. (2018). Use of principal component analysis (PCA) and hierarchical cluster analysis (HCA) for multivariate association between bioactive compounds and functional properties in foods: A critical perspective. *Trends in Food Science & Technology*, *72*, 83–90. https://doi.org/10.1016/j.tifs.2017.12.006

Graves, A., Mohamed, A. R., & Hinton, G. (2013). Speech recognition with deep recurrent neural networks. In *2013 IEEE International Conference on Acoustics, Speech and Signal Processing* (pp. 6645–6649). https://doi.org/10.1109/ICASSP.2013.6638947

Hassannejad, H., Matrella, G., Ciampolini, P., De Munari, I., Mordonini, M., & Cagnoni, S. (2016). Food image recognition using very deep convolutional networks. In *Proceedings of the 2nd International Workshop on Multimedia Assisted Dietary Management* (pp. 41–49). https://doi.org/10.1145/2986035.2986042

He, K. M., Zhang, X. Y., Ren, S. Q., & Sun, J. (2016). Deep residual learning for image recognition. In *2016 IEEE Conference on Computer Vision and Pattern Recognition* (pp. 770–778). https://doi.org/10.1109/CVPR.2016.90

Heravi, E. J., Aghdam, H. H., & Puig, D. (2015). A deep convolutional neural network for recognizing foods. In *Proceedings of 8th International Conference on Machine Vision* (Vol. 9875). https://doi.org/10.1117/12.2228875

Heravi, E. J., Aghdam, H. H., & Puig, D. (2017). Classification of foods by transferring knowledge from ImageNet dataset. In *Proceedings of 9th International Conference on Machine Vision* (Vol. 10341). https://doi.org/10.1117/12.2268737

Heravi, E. J., Aghdam, H. 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. https://doi.org/10.1016/j.patrec.2017.12.007

Herruzo, P., Bolaños, M., & Radeva, P. (2016). Can a cnn recognize Catalan diet? In *AIP Conference Proceedings* (Vol. 1773). https://doi.org/10.1063/1.4964956

Hochreiter, S., & Schmidhuber, J. (1997). Long short-term memory. *Neural Computation*, *9*(8), 1735–1780. https://doi.org/10.1162/neco.1997.9.8.1735

Jia, W., Li, Y., Qu, R., Baranowski, T., Burke, L. E., Zhang, H., . . . Mao, Z. H. (2018). Automatic food detection in egocentric images using artificial intelligence technology. *Public Health Nutrition*, *22*(7), 1168–1179. https://doi.org/10.1017/S1368980018000538

Jia, Y., Shelhamer, E., Donahue, J., Karayev, S., & Darrell, T. (2014). Caffe: Convolutional architecture for fast feature embedding. In *Proceedings of the 22nd ACM international conference on Multimedia* (pp. 675–678). https://doi.org/10.1145/2647868.2654889

Joutou, T., & Yanai, K. (2009). A food image recognition system with multiple mernel learning. In *2009 16th IEEE International Conference on*

Image Processing (ICIP) (pp. 285–288).
https://doi.org/10.1109/ICIP.2009.5413400

Jun, Y. H., Eo, T., Kim, T., Shin, H., Hwang, D., Bae, S. H., . . . Ahn, S. S. (2018). Deep-learned 3D black-blood imaging using automatic labelling technique and 3D convolutional neural networks for detecting metastatic brain tumors. *Scientific Reports*, 8. https://doi.org/10.1038/s41598-018-27742-1

Kamilaris, A., & Prenafeta-Boldu, F. X. (2018). Deep learning in agriculture: A survey. *Computers and Electronics in Agriculture*, *147*, 70–90. https://doi.org/10.1016/j.compag.2018.02.016

Kawano, Y., & Yanai, K. (2014). Food image recognition with deep convolutional features. In *Proceedings of the 2014 ACM International Joint Conference on Pervasive and Ubiquitous Computing: Adjunct Publication* (pp. 589-593). https://doi.org/10.1145/2638728.2641339

Kawano, Y., & Yanai, K. (2015). Automatic expansion of a food image dataset leveraging existing categories with domain adaptation. In L. Agapito, M. Bronstein, & . Rother (Eds.), *Computer Vision - Eccv 2014 Workshops* (Vol. 8927, pp. 3–17). https://doi.org/10.1007/978-3-319-16199-0_1

Kemker, R., Luu, R., & Kanan, C. (2018). Low-shot learning for the semantic segmentation of remote sensing imagery. *IEEE Transactions on Geoscience and Remote Sensing*, *56*(10), 6214–6223. https://doi.org/10.1109/tgrs.2018.2833808

Ketkar, N. (2014). Stochastic gradient descent. In F. Chollet (Ed.), *Deep learning with Python* (pp. 113–132). Berkeley, CA: Apress. https://doi.org/10.1007/978-1-4842-2766-4_8

Kim, Y. (2014). Convolutional neural networks for sentence classification. In *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing* (pp. 1746–1751). https://doi.org/10.3115/v1/D14-1181

Kingma, D., & Ba, J. (2014). *Adam: A method for stochastic optimization.* arXiv:1412.6980.

Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2012). ImageNet classification with deep convolutional neural networks. In *Proceedings of the 25th International Conference on Neural Information Processing Systems* (pp. 1097–1105).

LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. *Nature*, *521*(7553), 436–444. https://doi.org/10.1038/nature14539

Lee, M. C., Chiu, S. Y., & Chang, J. W. (2017). A deep convolutional neural network based Chinese menu recognition app. *Information Processing Letters*, *128*, 14–20. https://doi.org/10.1016/j.ipl.2017.07.010

Liu, C., Cao, Y., Luo, Y., Chen, G., Vokkarane, V., Ma, Y., . . . Hou, P. (2018). 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. https://doi.org/10.1109/tsc.2017.2662008

Liu, C., Cao, Y., Luo, Y., Chen, G. L., Vokkarane, V., & Ma, Y. S. (2016a). DeepFood: Deep learning-based food image recognition for computer-aided dietary assessment. In C. K. Chang, L. Chiari, Y. Cao, H. Jin, M. Mokhtari, & H. Aloulou (Eds.), *Inclusive Smart Cities and Digital Health* (Vol. 9677, pp. 37–48). https://doi.org/10.1007/978-3-319-39601-9_4

Liu, W., Anguelov, D., Erhan, D., Szegedy, C., Reed, S., Fu, C.-Y., & Berg, A. C. (2016b). SSD: Single shot multibox detector. In B. Leibe, J. Matas, N. Sebe, & M. Welling (Eds.), *Computer Vision - ECCV 2016, Pt I* (Vol. 9905, pp. 21–37). https://doi.org/10.1007/978-3-319-46448-0_2

Liu, Z., He, Y., Cen, H., & Lu, R. (2018). Deep feature representation with stacked sparse auto-encoder and convolutional neural network for hyperspectral imaging-based detection of cucumber defects. *Transactions of the ASABE*, *61*(2), 425–436. https://doi.org/10.13031/trans.12214

Luaces, O., Díez, Jorge, Joachims, T., & Bahamonde, A. (2015). Mapping preferences into euclidean space. *Expert Systems with Applications*, *42*(22), 8588–8596. https://doi.org/10.1016/j.eswa.2015.07.013

Lule, S. U., & Xia, W. S. (2005). Food phenolics, pros and cons: A review. *Food Reviews International*, *21*(4), 367–388. https://doi.org/10.1080/87559120500222862

Ma, B.-Q. (2017). Food packaging printing defect detection method based on image wavelet transform. *Food Research and Development*, *38*(5), 212–215. https://doi.org/10.3969/j.issn.1005-6521.2017.05.046

Mao, D. H., Wang, F., Hao, Z. H., & Li, H. S. (2018). Credit evaluation system based on blockchain for multiple stakeholders in the food supply chain. *International Journal of Environmental Research and Public Health*, *15*(8), 21. https://doi.org/10.3390/ijerph15081627

Martinel, N., Foresti, T. L., & Micheloni, C. (2018). Wide-slice residual networks for food recognition. In *2018 IEEE Winter Conference on*

Applications of Computer Vision (pp. 567–576). https://doi.org/10.1109/WACV.2018.00068

Matsuda, Y., Hoashi, H., & Yanai, K. (2012). Recognition of multiple-food images by detecting candidate regions, In *2012 IEEE International Conference on Multimedia and Expo* (pp. 25–30). https://doi.org/10.1109/ICME.2012.157

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. https://doi.org/10.1016/j.compbiomed.2018.02.008

Mezgec, S., & Seljak, B. K. (2017). NutriNet: A deep learning food and drink image recognition system for dietary assessment. *Nutrients*, *9*(7). https://doi.org/10.3390/nu9070657

Miotto, R., Wang, F., Wang, S., Jiang, X., & Dudley, J. T. (2018). Deep learning for healthcare: Review, opportunities and challenges. *Briefings in Bioinformatics*, *19*(6), 1236–1246. https://doi.org/10.1093/bib/bbx044

Mithun, B. S., Shinde, S., Bhavsar, K., Chowdhury, A., Mukhopadhyay, S., Gupta, K., . . . Kimbahune, S. (2018). Non-destructive method to detect artificially ripened banana using hyperspectral sensing and RGB imaging. In M. S. Kim, K. Chao, B. A. Chin, & B. K. Cho (Eds.), *Sensing for Agriculture and Food Quality and Safety X* (Vol. 10665). https://doi.org/10.1117/12.2306367

Monakhova, Y. B., Tsikin, A. M., Kuballa, T., Lachenmeier, D. W., & Mushtakova, S. P. (2014). Independent component analysis (ICA) algorithms for improved spectral deconvolution of overlapped signals in H-1 NMR analysis: Application to foods and related products. *Magnetic Resonance in Chemistry*, *52*(5), 231–240. https://doi.org/10.1002/mrc.4059

Myers, A., Johnston, N., Rathod, V., Korattikara, A., Gorban, A., Silberman, N., . . . Murphy, K. (2015). Im2Calories: Towards an automated mobile vision food diary. In *2015 IEEE International Conference on Computer Vision* (pp. 1233–1241). https://doi.org/10.1109/ICCV.2015.146

Naritomi, S., Tanno, R., Ege, T., & Yanai, K. (2018). FoodChangeLens: CNN-based food transformation on HoloLens. In *2018 IEEE International Conference on Artificial Intelligence and Virtual Reality (AIVR)* (pp. 197–199). https://doi.org/10.1109/AIVR.2018.00046

Ng, H. W., Nguyen, D., Vonikakis, V., & Winkler, S. (2015). Deep Learning for Emotion Recognition on Small Datasets using Transfer Learning. In *Proceedings of the 2015 ACM on International Conference on Multimodal Interaction (ICMI'15)*, 443–449. https://doi.org/10.1145/2818346.2830593

Noda, K., Yamaguchi, Y., Nakadai, K., Okuno, H. G., & Ogata, T. (2015). Audio-visual speech recognition using deep learning. *Applied Intelligence*, *42*(4), 722–737. https://doi.org/10.1007/s10489-014-0629-7

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. https://doi.org/10.1109/lsp.2017.2758862

Partel, V., Kakarla, C., & Ampatzidis, Y. (2019). Development and evaluation of a low-cost and smart technology for precision weed management utilizing artificial intelligence. *Computers and Electronics in Agriculture*, *157*, 339–350. https://doi.org/10.1016/j.compag.2018.12.048

Pérez-Palacios, T., Caballero, D., Antequera, T., Durán, M. L., Ávila, M., & Caro, A. (2017). Optimization of MRI acquisition and texture analysis to predict physico-chemical parameters of Loins by data mining. *Food and bioprocess technology*, *10*(4), 750–758. https://doi.org/10.1007/s11947-016-1853-4

Pfisterer, K. J., Amelard, R., Chung, A. G., & Wong, A. (2018). A new take on measuring relative nutritional density: The feasibility of using a deep neural network to assess commercially-prepared pureed food concentrations. *Journal of Food Engineering*, *223*, 220–235. https://doi.org/10.1016/j.jfoodeng.2017.10.016

Pierson, H. A., & Gashler, M. S. (2017). Deep learning in robotics: A review of recent research. *Advanced Robotics*, *31*(16), 821–835. https://doi.org/10.1080/01691864.2017.1365009

Pouladzadeh, P., Villalobos, G., Almaghrabi, R., & Shirmohammadi, S. (2012). A novel SVM based food recognition method for calorie measurement applications. In *2012 IEEE International Conference on Multimedia and Expo Workshops* (pp. 495–498). https://doi.org/10.1109/ICMEW.2012.92

Qannari, E. M. (2017). Sensometrics approaches in sensory and consumer research. *Current Opinion in Food Science*, *15*, 8–13. https://doi.org/10.1016/j.cofs.2017.04.001

Qiu, Z. J., Chen, J., Zhao, Y. Y., Zhu, S. 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), 12. https://doi.org/10.3390/app8020212

Ragusa, F., Tomaselli, V., Furnari, A., Battiato, S., & Farinella, G. M. (2016). Food vs non-food classification. *Proceedings of the 2nd International Workshop on Multimedia Assisted Dietary Management* (pp. 77–81). https://doi.org/10.1145/2986035.2986041

Ravikanth, L., Jayas, D. S., White, N. D. G., Fields, P. G., & Sun, D. W. (2017). Extraction of spectral information from hyperspectral data and application of hyperspectral imaging for food and agricultural products. *Food and Bioprocess Technology*, 10(1), 1–33. https://doi.org/10.1007/s11947-016-1817-8

Ren, S., He, K., Girshick, R., & Sun, J. (2017). Faster R-CNN: Towards real-time object detection with region proposal networks. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 39(6), 1137–1149. https://doi.org/10.1109/tpami.2016.2577031

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. https://doi.org/10.1007/s13748-017-0137-1

Ropodi, A. I., Panagou, E. Z., & Nychas, G. J. E. (2016). Data mining derived from food analyses using non-invasive/non-destructive analytical techniques; determination of food authenticity, quality & safety in tandem with computer science disciplines. *Trends in Food Science & Technology*, 50, 11–25. https://doi.org/10.1016/j.tifs.2016.01.011

Schmidhuber, J. (2015). Deep learning in neural networks: An overview. *Neural Networks*, 61, 85–117. https://doi.org/10.1016/j.neunet.2014.09.003

Shelhamer, E., Long, J., & Darrell, T. (2017). Fully convolutional networks for semantic segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 39(4), 640–651. https://doi.org/10.1109/tpami.2016.2572683

Shen, D., Wu, G., & Suk, H.-I. (2017). Deep learning in medical image analysis. In M. L. Yarmush (Ed.), *Annual review of biomedical engineering* (Vol. 19, pp. 221–248). Palo Alto, CA: Annual Reviews. https://doi.org/10.1146/annurev-bioeng-071516-044442

Simonyan, K., & Zisserman, A. (2014). Very deep convolutional networks for large-scale image recognition. Retrieved from http://arxiv.org/abs/1409.1556

Singla, A., Yuan, L., & Ebrahimi, T. (2016). Food/non-food image classification and food categorization using pre-trained GoogLeNet model. In *Proceedings of the 2nd International Workshop on Multimedia Assisted Dietary Management* (pp. 3–11). https://doi.org/10.1145/2986035.2986039

Song, Q., Zheng, Y. J., Xue, Y., Sheng, W. G., & Zhao, M. R. (2017). An evolutionary deep neural network for predicting morbidity of gastrointestinal infections by food contamination. *Neurocomputing*, 226, 16–22. https://doi.org/10.1016/j.neucom.2016.11.018

Stelick, A., & Dando, R. (2018). Thinking outside the booth—The eating environment, context and ecological validity in sensory and consumer research. *Current Opinion in Food Science*, 21, 26–31. https://doi.org/10.1016/j.cofs.2018.05.005

Suk, H. I., Lee, S. W., Shen, D. G., & Alzheimer's Dis, N. (2015). Latent feature representation with stacked auto-encoder for AD/MCI diagnosis. *Brain Structure & Function*, 220(2), 841–859. https://doi.org/10.1007/s00429-013-0687-3

Sun, Y., Wei, K. L., Liu, Q., Pan, L. Q., & Tu, K. (2018). Classification and discrimination of different fungal diseases of three infection levels on peaches using hyperspectral reflectance imaging analysis. *Sensors*, 18(4). https://doi.org/10.3390/s18041295

Sutskever, I., Vinyals, O., & Le, Q. V. (2014). Sequence to sequence learning with neural networks. In Z. Ghahramani, M. Welling, C. Cortes, N. D. Lawrence, & K. Q. Weinberger (Eds.), *Advances in Neural Information Processing Systems* (Vol. 27). Retrieved from https://arxiv.org/abs/1409.3215

Szegedy, C., Liu, W., Jia, Y. Q., Sermanet, P., Reed, S., Anguelov, D., & Rabinovich, A. (2015). Going deeper with convolutions. In *2015 IEEE Conference on Computer Vision and Pattern Recognition* (pp. 1–9). https://doi.org/10.1109/CVPR.2015.7298594

Tan, W. X., Zhao, C. J., & Wu, H. R. (2016). Intelligent alerting for fruit-melon lesion image based on momentum deep learning. *Multimedia Tools and Applications*, 75(24), 16741–16761. https://doi.org/10.1007/s11042-015-2940-7

Tatsuma, A., & Aono, M. (2016). Food image recognition using covariance of convolutional layer feature maps. *IEICE Transactions on Information and Systems*, E99D(6), 1711–1715. https://doi.org/10.1587/transinf.2015EDL8212

Tian, H., Li, F., Qin, L., Yu, H., & Ma, X. (2014). Discrimination of chicken seasonings and beef seasonings using electronic nose and sensory evaluation. *Journal of Food Science*, 79(11), 2346–2353. https://doi.org/10.1111/1750-3841.12675

Vedaldi, A., & Lenc, K. (2015). MatConvNet convolutional neural networks for MATLAB. In *Proceedings of the 23rd ACM International Conference on Multimedia* (pp. 689–692). https://doi.org/10.1145/2733373.2807412

Wang, T., Chen, J., Fan, Y., Qiu, Z., & He, Y. (2018a). SeeFruits: Design and evaluation of a cloud-based ultra-portable NIRS system for sweet cherry quality detection. *Computers and Electronics in Agriculture*, 152, 302–313. https://doi.org/10.1016/j.compag.2018.07.017

Wang, Z. D., Hu, M. H., & Zhai, G. T. (2018b). Application of deep learning architectures for accurate and rapid detection of internal mechanical damage of blueberry using hyperspectral transmittance data. *Sensors*, 18(4). https://doi.org/10.3390/s18041126

Wu, D., Sharma, N., & Blumenstein, M. (2017). Recent advances in video-based human action recognition using deep learning: A review. In *2017 International Joint Conference on Neural Networks* (pp. 2865–2872). https://doi.org/10.1109/IJCNN.2017.7966210

Wu, H., Merler, M., Uceda-Sosa, R., & Smith, J. R. (2016). Learning to make better mistakes: Semantics-aware visual food recognition. In *Proceedings of the 24th ACM International Conference on Multimedia* (pp. 172–176). https://doi.org/10.1145/2964284.2967205

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). https://doi.org/10.3390/molecules23112831

Yanai, K., & Kawano, Y. (2015). Food image recognition using deep convolutional neural network with pre-training and fine-tuning. In *2015 IEEE International Conference on Multimedia & Expo Workshops* (pp. 1–6). https://doi.org/10.1109/ICMEW.2015.7169816

Yiqun, H., Kangas, L. J., & Rasco, B. A. (2007). Applications of artificial neural networks (ANNs) in food science. *Critical Reviews in Food Science and Nutrition*, 47(2), 113–126. https://doi.org/10.1080/10408390600626453

Yordi, E. G., Koelig, R., Mota, Y. C., Matos, M. J., Santana, L., Uriarte, E., & Molina, E. (2015). Application of KNN algorithm in determining the total antioxidant capacity of flavonoid-containing foods. In *19th International Electronic Conference on Synthetic Organic Chemistry*. https://doi.org/10.3390/ecsoc-19-e002

Yu, X. J., Lu, H. D., & Wu, D. (2018). Development of deep learning method for predicting firmness and soluble solid content of postharvest Korla fragrant pear using Vis/NIR hyperspectral reflectance imaging. *Postharvest Biology and Technology*, 141, 39–49. https://doi.org/10.1016/j.postharvbio.2018.02.013

Yu, X. J., Tang, L., Wu, X. F., & Lu, H. D. (2018). Nondestructive freshness discriminating of shrimp using visible/near-infrared hyperspectral imaging technique and deep learning algorithm. *Food Analytical Methods*, 11(3), 768–780. https://doi.org/10.1007/s12161-017-1050-8

Yu, X. J., Wang, J. P., Wen, S. T., Yang, J. Q., & Zhang, F. F. (2019). A deep learning based feature extraction method on hyperspectral images for nondestructive prediction of TVB–N content in Pacific white shrimp (*Litopenaeus vannamei*). *Biosystems Engineering*, 178, 244–255. https://doi.org/10.1016/j.biosystemseng.2018.11.018

Yuan, Q. Q., Zhang, Q., Li, J., Shen, H. F., & Zhang, L. P. (2019). Hyperspectral image denoising employing a spatial-spectral deep residual convolutional neural network. *IEEE Transactions on Geoscience and Remote Sensing*, 57(2), 1205–1218. https://doi.org/10.1109/tgrs.2018.2865197

Zhang, W. S., Zhang, Y. J., Zhai, J., Zhao, D. H., Xu, L., Zhou, J. H., . . . Yang, S. (2018a). Multi-source data fusion using deep learning for smart refrigerators. *Computers in Industry*, 95, 15–21. https://doi.org/10.1016/j.compind.2017.09.001

Zhang, X. R., Sun, Y. J., Zhang, J. Y., Wu, P., & Jiao, L. C. (2018b). Hyperspectral unmixing via deep convolutional neural networks. *IEEE Geoscience and Remote Sensing Letters*, 15(11), 1755–1759. https://doi.org/10.1109/lgrs.2018.2857804

Zheng, J., Zou, L., & Wang, Z. J. (2018). Mid-level deep food part mining for food image recognition. *IET Computer Vision*, 12(3), 298–304. https://doi.org/10.1049/iet-cvi.2016.0335

Appendix A–Applications of Deep Learning in Quality Detection of Fruit and Vegetables

Objective	Dataset and data type	Labels	Data preprocessing	Data augmentation	Frameworks (or tools) and networks	Performance metric and metric value	Comparison
Fruit recognition in smart refrigerator (Zhang et al., 2018a)	RGB images: Fruit picture resources on the Internet and images set created by authors	10 classes:10 different kinds of fruits	/	/	*Caffe* (Y. Jia, Shelhamer, Donahue, Karayev, & Darrell, 2014) multimodel fusion architecture based on SSD model (Resnet, VGG16, VGG19)	Classification Accuracy: 97%	SSD (ResNet): 91% SSD (VGG16): 89% SSD (VGG19): 90% Multimodel fusion: 92%
Prediction of firmness and SSC of Korla fragrant pear (Yu et al., 2018)	Hyperspectral data: Place 180 pear fruit in chambers. Take out 15 pears every other day to obtain spectral data, and measure their reference.	Firmness and SSC measured by destructive methods.	calculate the mean spectra in ROI as the input of the SAE-FNN model	/	*Keras* *https://keras.io SAE and FNN	$R_P{}^2$: 0.890, RMSEP: 1.81N, RPD$_P$: 3.05 for firmness, and $R_P{}^2$: 0.921, RMSEP: 0.22%, RPD$_P$: 3.68 for SSC.	Better than using PLSR and LS-SVM methods
Detection of artificially ripened banana (Mithun et al., 2018)	RGB images: 176 naturally ripened bananas and 151 artificially ripened bananas	Two classes "naturally ripened" and "artificially ripened"	Resize images to 512*512 resolution	/	*Caffe* Modified AlexNet architecture	Classification Accuracy: 90%	/
Mangosteen surface defect detection (Azizah et al., 2017)	Authors collected: 120 RGB images of mangosteen (30 defect images and 90 fine images)	"Defect" and "Fine"	Crop and resize the images to 512*512 pixels	/	/ A simple CNN defined by authors	Classification Accuracy: 97.5%	/
Plum varieties classification (Rodriguez et al., 2018)	Authors collected: 525 RGB images of plum with a wide range of recollection dates	Three classes: "Black Splendor," "OwenT" and "Angelino"	Move the background using image segmentation method	/	*Caffe* AlexNet	Classification Accuracy: 91% to 97%	/

Appendix A—Continued

Objective	Dataset and data type	Labels	Data preprocessing	Data augmentation	Frameworks (or tools) and networks	Performance metric and metric value	Comparison
Internal mechanical damage of blueberry detection (Wang et al., 2018b)	Hyperspectral data collected by authors: 557 blueberries	Two classes: damage and sound	The size of each sample was converted to 32*32 with 151 channels by clipping, segmentation, resizing, and subsampling methods	Expand the training images by flipping vertically, and horizontally, rotating, and randomly cropping	*TensorFlow* (Abadi, Agarwal, Barham, Brevdo, & Zheng, 2016) ResNet and ResNeXt	Average accuracy and F1-score: ResNet (88.44%/0.8784) ResNeXt (89.52%/0.8905)	Conventional classifier: SMO (80.82%/0.8268), LR (76.06%/0.7796), RF (0.7314/0.7529), Bagging (0.7113/0.7339), MLP (0.7827/0.7971)
Discrimination of fungal diseases on peaches (Sun et al., 2018)	Hyperspectral data: four groups of peaches (30 for the control group, 270 for the three treated groups)	Three classes: diseased and healthy peaches under three levels of decay. Four classes: peaches with three fungal diseases and healthy samples	Combine features from spectral data and image features. Obtain the image of the first principal component using PCA.	/	*Matlab* (Vedaldi & Lenc, 2015) DBN	Classification Accuracy: 85% to 100% (classification of diseased and healthy peaches); 60% to 100% (classification of three fungal diseases)	The DBN models achieve higher classification accuracy at the slightly decayed level compared to the PLSDA model.
Fruit skin lesion recognition (Tan, Zhao, & Wu, 2016)	Approximately 250 infrared image samples of four kinds of diseased apples (scab skin, black rot, scar skin, and ring spot)	Four classes: four different diseased	Reduce dimension using PCA method	Intensity adjustment and rotation translation	*Matlab* (Vedaldi & Lenc, 2015) A five-layer CNN model defined by authors	Classification Accuracy: 97.5 %	Multilayer perceptron (68.75%) and k-NN (62.50%).
Cucumbers defect detection (C. Liu et al., 2018)	Hyperspectral data: 230 samples with five classes	Five classes: normal, watery, split/hollow, shrivel, and surface defect	Flat field correction and image segmentation	/	*Caffe* Stacked sparse auto-encoder combined with convolutional neural network (CNN-SSAE)	Classification Accuracy with conveyor Speed of 85 mm/s & 165 mm/s: 91.1% / 88.3%	extended morphological profile-SVM: 68.3%/60.8% Bag-of-visual-words: 73.0% / 62.2%

Appendix B–The Best Performances Achieved by Deep Learning Methods on Different Public Food Image Databases

Databases	URL of databases	Best performance Top-1% & Top-5%[a]	Network	Reference
Food-101 (Bossard et al., 2014)	http://www.vision.ee.ethz.ch/datasets/food-101/.	90.27, 98.71	WISeR	Martinel et al. (2018)
UECFood-256 (Kawano & Yanai, 2015)	http://foodcam.mobi/dataset/	83.15, 95.45	WISeR	Martinel et al. (2018)
UECFood-100 (Matsuda et al., 2012)	http://foodcam.mobi/dataset/	89.58, 99.23	WISeR	Martinel et al. (2018)
Food-50 (Joutou & Yanai, 2009)	*not mentioned in the original paper	93.84, 99.44	ResNet-50	Ciocca et al. (2018)
Food-524 (Ciocca et al., 2017)	http://www.ivl.disco.unimib.it/activities/food524db/	81.34, 95.45	ResNet-50	Ciocca et al. (2017)
Food-475 (Ciocca et al., 2018)	http://www.ivl.disco.unimib.it/activities/food475db/	81.59, 95.50	ResNet-50	Ciocca et al. (2018)
VIREO Food172 (Chen & Ngo, 2016)	http://vireo.cs.cityu.edu.hk/VireoFood172/	85.86, 97.32	ResNet-50	Ciocca et al. (2018)

[a] Top-1% & Top-5% denote the Top-1 and Top-5 classification accuracy for evaluation.