Supplementary Materials

Deep Learning in Food Image Recognition: A Comprehensive Review

Detianjun Liu[1], Enguang Zuo[1,*], Dingding Wang[1,*], Liang He[1,2,3], Liujing Dong[1], Xinyao Lu[1]

[1] School of Intelligence Science and Technology, Xinjiang University, Urumqi 830017, China
[2] Xinjiang Key Laboratory of Signal Detection and Processing, Urumqi 830017, China
[3] Department of Electronic Engineering, and Beijing National Research Center for information Science and Technology Tsinghua University, Beijing 10084, China

[*] Correspondence: wl_826@xju.edu.cn(Dingding Wang)

Table S1 Summary of main methods

Reference	Year	Main Focus	Best Evaluation Performance
Parrish et al.**Error! Reference source not found.**	1977	Pictorial pattern recognition applied to fruit harvesting	/
Bolle et al.**Error! Reference source not found.**	1996	Automatic product 1D system: VeggieVision	ACC: 95%
Kitamura et al.**Error! Reference source not found.**	2008	Food-logging system	/
Pornpanomchai et al.**Error! Reference source not found.**	2009	Thai fruit recognition system (TFRS)	/
Wu et al.**Error! Reference source not found.**	2009	Identify food directly from video	/
Yang et al.**Error! Reference source not found.**	2010	A soft pixel-level segmentation of the image	/
Kong et al.**Error! Reference source not found.**	2011	An application: DietCam	/
Matsuda et al.**Error! Reference source not found.**	2012	A two-step method to recognize multi-countriesple-food images	Classification rate: 55.8%
Chen et al.**Error! Reference source not found.**	2012	Automatic Chinese food identification and quantity estimation	ACC: 68.3%
Wang et al.**Error! Reference source not found.**	2013	Pervasive cloud computing paradigm to improve the performance of food image recognition	/
Kagaya et al.**Error! Reference source not found.**	2014	Convolutional neural network (CNN)	/
Kawano et al.**Error! Reference source not found.**	2014	A mobile food recognition system: FoodCam	Classification rate: 79.2 %
Anthimopoulos et al.**Error! Reference source not found.**	2014	A bag-of-features (BoF) model	ACC: 78%
Song et al.**Error! Reference source not found.**	2014	Semantic features learning and location-adaptive classification	/
Tammachat et al.**Error! Reference source not found.**	2014	Recognize images of food taken by users	/
Yanai et al.**Error! Reference source not found.**	2015	Deep convolutional neural network (DCNN)	Top-1 ACC: 78.77%
Herranz et al.**Error! Reference source not found.**	2015	Contextual knowledge	/

Ahmed et al.**Error! Reference source not found.**	2015	A bag-of-features (BoF) based on SURF bag of SURF features bag-of HOG Features	ACC: 72%
Myers et al.**Error! Reference source not found.**	2015	CNN-based classifier A system: Im2Calories	/
Yi et al.**Error! Reference source not found.**	2015	A wearable acoustic sensor system :AutoDietary	ACC: 84.9%
Singla et al.**Error! Reference source not found.**	2016	A GoogLeNet model based on deep convolutional neural network	Food/Non-food classification: 99.2% Food category recognition: 83.6%
Liu et al.**Error! Reference source not found.**	2016	Convolutional Neural Network	Top-1 ACC: 77.4% Top-5 ACC: 93.7%
Zheng et al.**Error! Reference source not found.**	2016	Superpixel-based image recognition	/
Tatsuma et al.**Error! Reference source not found.**	2016	Covariances of convolutional layer feature maps	ACC: 58.65%
Mezgec et al.**Error! Reference source not found.**	2017	NutriNet	Classification ACC: 86.72% Detection dataset ACC: 94.47%
Pouladzadeh et al.**Error! Reference source not found.**	2017	A mobile food recognition system	Recall rate: 90.98%, Precision rate: 93.05% ACC: 94.11%
Giovany et al.**Error! Reference source not found.**	2017	SIFT , Machine learning	/
Termritthikun et al.**Error! Reference source not found.**	2017	NU-InNet (nararesuan University Inception network)	/
Zheng et al.**Error! Reference source not found.**	2017	Linear Distance Coding (LDC) framework	Top-1 ACC: 83.07% Top-5 ACC: 97.04%
Termritthikun et al.**Error! Reference source not found.**	2018	NU-ResNet	/
Horiguchi et al.**Error! Reference source not found.**	2018	Nearest class mean classifier,1-nearest neighbor classifier, Deep features	/

Ming et al.**Error! Reference source not found.**	2018	Food Photo Recognition for Dietary Tracking	/
Reddy et al.**Error! Reference source not found.**	2019	A calorie measurement system	Testing ACC: 78.7% Training ACC: 93.29%
Sahoo et al.**Error! Reference source not found.**	2019	A smart-food logging system: FoodAI	/
Jiang et al.**Error! Reference source not found.**	2019	A multi-countries-scale multi-countries-view feature aggregation (MSMVFA) scheme	/
Merchant et al.**Error! Reference source not found.**	2019	A CNN-based food recognition mobile application:ConvFood	/
Jeny et al.**Error! Reference source not found.**	2019	FoNet - Local food recognition Deep residual neural networks	Highest ACC: 98.16%
Tasci et al.**Error! Reference source not found.**	2020	Voting combinations-based ensemble of fine-tuned convolutional neural networks	/
Metwalli et al.**Error! Reference source not found.**	2020	A DenseFood model based on densely connected convolutional neural network	ACC: 81.23%
Song et al.**Error! Reference source not found.**	2020	Hybrid attention，Convolutional neural network (CNN)	/
He et al.**Error! Reference source not found.**	2020	An end-to-end multi-countries-task framework	/
Jiang et al.**Error! Reference source not found.**	2020	A multi-countries-View Few-Shot Learning (MVFSL) framework	/
Zhao et al.**Error! Reference source not found.**	2020	Joint-learning Distilled Network（JDNet）	UECFood-256: 84.0% Food-101: 91.2%
Minija et al.**Error! Reference source not found.**	2020	multi-countriesple-hypotheses image segmentation Feed-forward neural network classifier	MAA: 0.947 SA: 0.959
Liang et al.**Error! Reference source not found.**	2020	A novel multi-countries-View Attention Network,multi-countries-task learning framework, incorporates multi-countriesple semantic features	/
He et al.**Error! Reference source not found.**	2021	YOLOv3, Convolution residual module, CBAM (Convolutional Block Attention Module) attention	/
Liu et al.**Error! Reference source not found.**	2021	Attention Fusion Network (AFN), Food-Ingredient Joint Learning module	/
Xiao et al.**Error! Reference source not found.**	2021	Jumping convolution, convolutional neural network (CNN)	/
Wang et al.**Error! Reference source not found.**	2021	A two-stage learning mode of YOLO-SIMM, two versions:YOLO-SiamV1 and YOLO-SiamV2	/

Liu et al.**Error! Reference source not found.**	2022	multi-countries-scale jigsaw and reconstruction network (MJR-Net)	ETH Food-101: 90.82% Vireo Food-172: 91.37% ISIA Food-500: 64.95%
Zhang et al.**Error! Reference source not found.**	2022	A non-wearable food recognition and weight estimation system (nWFWS)	ACC:89.60%
Sheng et al.**Error! Reference source not found.**	2022	A Lightweight Transformer-Based Deep Neural Network	/
Liu et al.**Error! Reference source not found.**	2022	EfficientDet deep learning (DL) model	mAP:0.92
Haque et al.**Error! Reference source not found.**	2022	Parameter-optimized Convolution Neural Networks (CNN)	/
Cherpanath et al.**Error! Reference source not found.**	2023	Faster R-CNN, Mask R-CNN	/
Luo et al.**Error! Reference source not found.**	2023	YOLOv5 convolutional neural network	/
Zuo et al.**Error! Reference source not found.**	2023	Predictive REcurrent Neural Network (PReNet), PyTorch	/
Liu et al.**Error! Reference source not found.**	2023	EfficientNet, ResNet	/
Ramkumar et al.**Error! Reference source not found.**	2023	Layer-based methodology, Intelligent Deep Learning	ACC: 98%
Zhu et al.**Error! Reference source not found.**	2023	Generative self-supervised learning	ACC: 85.99%
Nadeem et al.**Error! Reference source not found.**	2023	A smartphone-based diet-tracking application:Smart Diet Diary	ACC: 80.1%
Jiang et al.**Error! Reference source not found.**	2024	Self-attention, Local attention mechanism	Top-1 ACC:65.58% Top-5 ACC:90.03%
Rokhva et al.**Error! Reference source not found.**	2024	MobileNetV2	ACC: 92.97%
Sheng et al.**Error! Reference source not found.**	2024	Global Shuffle Net (GSNet)	Top-1 ACC: 87.9%

Phiphitphatphaisit et al.**Error! Reference source not found.**	2024	ASTFF-Net,ResNet50	/
Xiao et al.**Error! Reference source not found.**	2024	Swin Transformer An end-to-end fine-grained food universal classifier A deep convolution module for obtaining local enhanced feature representation and combines it with the global feature	/
Boyd et al.**Error! Reference source not found.**	2024	A mobile application: FridgeSnap	Training accuracy: 0.99 Training loss: 0.01 Validation accuracy: 0.79 Validation loss: 0.92
Abiyev et al.**Error! Reference source not found.**	2024	A robust food recognition model using deep convolutional neural networks with a self-attention mechanism (FRCNNSAM)	Food-101: 96.40% MA Food-121: 95.11%
Bu et al.**Error! Reference source not found.**	2024	Transfer learning, Ensemble learning	ACC: 96.88%
Sheng et al.**Error! Reference source not found.**	2024	An Efficient Hybrid Food Recognition Net (EHFR–Net) Vision Transformer (ViT)	ACC: 90.7%
Wang et al.**Error! Reference source not found.**	2024	EfficientNet, Swin Transformer Feature Pyramid Network (FPN)	Nutrition5k: 79.50% ChinaMartFood109: 80.25%
Liu et al.**Error! Reference source not found.**	2024	Convolution-Enhanced Bi-Branch Adaptive Transformer (CBiAFormer) Structural Learning and Cross-Task Interaction (SLCI) Bi-Branch Adaptive Attention (BiA-Attention)	/

Table S2 Literature Selection Criteria

Category	Criteria
Databases	Web of Science, IEEE Xplore, PubMed, ACM DL, CNKI, Arxiv
Keywords	"food image recognition" + "deep learning" + ("dataset" OR "CNN" OR "Transformer")
Timeframe	2014–2024
Inclusion	Empirical validation, public datasets, DL-based methods
Exclusion	Non-DL approaches, non-reproducible studies

Table S3 Summary of datasets

Datasets	Year	Class	Images	Type	Public
PFID**Error! Reference source not found.**	2009	101	4545	Western	×
Food50**Error! Reference source not found.**	2009	50	5000	Misc.	×
Food85**Error! Reference source not found.**	2010	85	8500	Misc.	×
UEC Food100**Error! Reference source not found.**	2012	100	14361	Japanese	√
UEC Food256**Error! Reference source not found.**	2014	256	25088	Japanese	√
ETHZ Food-101**Error! Reference source not found.**	2014	101	101000	Western	√
Diabetes**Error! Reference source not found.**	2014	11	4868	Misc.	×
UPMC Food-101**Error! Reference source not found.**	2015	101	90840	Western	√
Geolocation-food**Error! Reference source not found.**	2015	/	117504	Chinese	√
UNIMIB2015**Error! Reference source not found.**	2015	15	2000	Misc.	√
UNICT-FD889**Error! Reference source not found.**	2015	889	3583	Misc.	√
UNIMIB2016**Error! Reference source not found.**	2016	73	1027	Misc.	√
Vireo Food-172**Error! Reference source not found.**	2016	172	110241	Chinese	√
Food-975**Error! Reference source not found.**	2016	975	37785	Misc.	×
Food500**Error! Reference source not found.**	2016	508	148408	Misc.	×
Food11**Error! Reference source not found.**	2016	11	16643	Misc.	×
UNICT-FD1200**Error! Reference source not found.**	2016	1200	4754	Misc.	√
Food524DB**Error! Reference source not found.**	2017	524	247636	Misc.	√
ChineseFoodNet**Error! Reference source not found.**	2017	208	192000	Chinese	√
VegFru**Error! Reference source not found.**	2017	292	160000	Misc.	√
FruitVeg-81**Error! Reference source not found.**	2017	81	15737	Misc.	√
Food-101**Error! Reference source not found.**03]	2017	101	101000	Misc.	√
Food-101N**Error! Reference source not found.**	2017	101	310009	Misc.	√
Yummly-28K**Error! Reference source not found.**	2017	29	27638	Western	√
Yummly-66K**Error! Reference source not found.**	2018	2440	66615	Western	√
ISIA Food-200**Error! Reference source not found.**	2019	200	197323	Misc.	√
Fruits 360**Error! Reference source not found.**	2019	131	90483	Misc.	√

Sushi-50**Error! Reference source not found.**	2019	50	3963	Japanese	√
FoodX-251**Error! Reference source not found.**	2019	251	158846	Misc.	√
FoodAI-756**Error! Reference source not found.**	2019	756	400000	Misc.	×
Grocery**Error! Reference source not found.**	2019	33	5125	Misc.	√
Refrigerator-Food dataset**Error! Reference source not found.**	2019	80	49557	Misc.	×
ISIA Food-500**Error! Reference source not found.**	2020	500	399726	Misc.	√
RGBD-FG**Error! Reference source not found.**	2020	50	93051	Misc.	×
Vireo Food-251**Error! Reference source not found.**	2020	251	169673	Chinese	√
Food2K**Error! Reference source not found.**	2021	2000	1036564	Misc.	√
FoodSeg103**Error! Reference source not found.**	2021	104	7118	Misc.	√
FoodLogoDet-1500**Error! Reference source not found.**	2021	1500	100000	Misc.	√
CNFOOD-241**Error! Reference source not found.**	2023	241	190000	Chinese	√
CNFOOD-241-Chen**Error! Reference source not found.**	2024	241	191811	Chinese	×

Table S4 Comparison of Lightweight Neural Network Models

Model	Size (MB)	FLOPs (B)	Top-1 Accuracy (%)
MobileNetV2[117]	3.4	0.3	72
MobileNetV2 (1.4)[117]	6.9	0.585	74.7
MobileNetV3-Large[37]	5.4	0.129	75.2
MobileNetV3-Small[37]	2.9	0.066	67.4
EfficientNet-B0[38]	5.3	0.39	77.1
EfficientNet-B1[38]	7.8	0.7	79.1
EfficientNet-B2[38]	9.2	1.0	80.1
EfficientNet-B3[38]	12	1.8	81.6
EfficientNet-B4[38]	19	4.2	82.9
EfficientNet-B5[38]	30	9.9	83.6
EfficientNet-B6[38]	43	19	84.0
EfficientNet-B7[38]	66	37	84.3
ShuffleNet v2 0.5×[118]	1.4	0.041	60.3
ShuffleNet v2 1×[118]	2.3	0.146	69.4
ShuffleNet v2 1.5×[118]	3.5	0.299	72.6
ShuffleNet v2 2×[118]	7.4	0.591	74.9
GSNet-1.0[14]	0.9	0.295	87.0
GSNet-1.0[14]	0.9	0.295	87.8
GSNet-1.0[14]	1.1	0.295	69.6
GSNet-1.0[14]	1.4	0.295	62.0
GSNet-1.5[14]	1.8	0.665	87.9
GSNet-1.5[14]	2.0	0.665	89.1
GSNet-1.5[14]	2.1	0.665	71.0
GSNet-1.5[14]	2.6	0.665	64.3
GSNet-2.0[14]	3.1	1.051	88.4
GSNet-2.0[14]	3.2	1.051	89.3
GSNet-2.0[14]	3.5	1.051	71.9
GSNet-2.0[14]	4.1	1.051	64.9

Table S5 Comparison of Deep Learning Models in Food Image Recognition

Model	Dataset Evaluated	Performance Metrics	Key Features
FRCNNSAM [22]	Food-101 MA Food-121	Accuracy: 96.40%(Food-101), 95.11%(MA Food-121)	Self-attention, CNN ensemble
Ensemble Model[23]	Food-11	Accuracy: 96.88%	Transfer learning , feature fusion
ASTFF-Net [24]	Food-11 UEC Food-100 UEC Food-256 ETH Food-101	Not Available	Spatiotemporal fusion, LSTM-enhanced
ViT-based[10]	ISIA Food-500	Top-1:65.58%, Top-5: 90.03%	Self-supervised pre-training
CBiAFormer and SLCI[25]	ETH Food-101 Vireo Food-172 ISIA Food-200	ETH Food-101:Top-1:92.40%, Top-5:98.87% Vireo Food-172:Top-1:91.58%, Top-5:98.75% ISIA Food-200:Top-1:70.92%, Top-5:93.21%	Transformer, multi-countries-task learning