Computer Vision for Dietary Assessment

Chia-Fang Chung
cfchung@iu.edu
Indiana University Bloomington
Bloomington, Indiana, USA

Alejandra Ramos
Case Western Reserve University
USA
axr738@case.edu

Pei-Ni Chiang
Indiana University Bloomington
Bloomington, Indiana, USA
pechia@iu.edu

Chien-Chun Wu
Indiana University Bloomington
Bloomington, Indiana, USA
chiewu@iu.edu

Connie Ann Tan
Indiana University Bloomington
Bloomington, Indiana, USA
cotan@iu.edu

Weslie Khoo
Indiana University Bloomington
Bloomington, Indiana, USA
weskhoo@iu.edu

David Crandall
Indiana University Bloomington
Bloomington, Indiana, USA
djcran@iu.edu

ABSTRACT

Automated visual recognition of food from smartphone cameras could be a powerful tool for assisting people to track their eating behaviors. Existing work in computer vision has focused on coarse-grained food classification, typically on idealized food images collected from the web, which may not reflect the challenges of real-world foods or photos. Despite advancements in computer vision over the last few years, error rates in these food recognition studies are quite high compared to human observers. We argue that we need to rethink how computer vision and AI can automate food logging, such as understanding the types of relationships humans have with foods, or creating semi-automatic tools that could complement dietitians instead of replacing them.

KEYWORDS

Dietary assessment; food recognition; computer vision; artificial intelligence

ACM Reference Format:
Chia-Fang Chung, Alejandra Ramos, Pei-Ni Chiang, Chien-Chun Wu, Connie Ann Tan, Weslie Khoo, and David Crandall. 2021. Computer Vision for Dietary Assessment. In *Proceedings of CHI Workshop on Realizing AI in Healthcare: Challenges Appearing in the Wild.* ACM, New York, NY, USA, 4 pages. https://doi.org/10.1145/nnnnnnn.nnnnnnn

1 INTRODUCTION

Empowering people to make good health choices begins by creating awareness of their current behaviors. Consumer smartphones and smartwatches have provided new tools for collecting these types of data, allowing people to monitor their physical activity, heart rate, sleep quality, blood glucose, etc. Mobile devices could also help people monitor their food choices, by having people quickly photograph meals and then using computer vision to automatically identify relevant dietary information. Taking food photos not only reduces the burden of keeping food diaries [9] but also provides social support in the pursuit of healthy eating goals when shared on social media [7]. In addition, food photos contain contextual information that can be useful for health experts to provide individualized diagnosis and treatment recommendations [25]. Computer vision-based technologies could provide immediate assessments to support between-visit recommendations, or to help individuals who do not have access to expert resources [8].

Despite progress in automatic food recognition in the computer vision community and a number of commercially-available smartphone applications that utilize this technology, automatic food logging has not become nearly as popular as fitness trackers or other health-related devices [2, 9]. Part of the problem may be that automatic food recognition is not accurate enough in the real world — which may be caused by a number of issues including imperfect computer vision algorithms, unrealistic training datasets, and inherent limitations in visual observation as a means for accurately estimating dietary content — or does not solve the types of problems that are most useful to users.

In this position paper, we briefly summarize recent work related to computer vision-based food recognition through the lens of applicability for real-world dietary assessment. Then, using data collected from a preliminary, empirical study, we contrast these computer vision approaches with review processes conducted by dietitians. Finally, we propose how limitations of current technology could be overcome or mitigated, such as by moving away from trying to recognize individual dishes and moving towards providing feedback on eating behaviors over time, or by creating semi-automatic tools that try to complement dietitians instead of replacing them.

CHI Workshop on Realizing AI in Healthcare: Challenges Appearing in the Wild, Realizing AI in HealthCare, May 8-9, 2021

ACM ISBN 978-x-xxxx-xxxx-x/YY/MM…$15.00
https://doi.org/10.1145/nnnnnnn.nnnnnnn

2 COMPUTER VISION-BASED FOOD RECOGNITION

Image recognition technology has seen tremendous progress over the last decade, driven in large part by advances in deep machine learning [26]. Most work in image recognition involves defining a discrete set of categories to be recognized (such as objects or scene types), collecting a large-scale image dataset of examples of each category (typically thousands of images), and training a machine learning model such as a Convolutional Neural Network (CNN) [23]. Unlike earlier approaches to computer vision, CNNs learn visual features directly from images, avoiding the need for programmers to create custom feature extraction algorithms for each new application.

Much work has studied visual recognition of food images. Here we give some examples of the major themes of research (see [19] for a comprehensive survey). Most work has been conducted by computer vision researchers interested in testing their models on new applications, and thus follows the same general classification paradigm. Bossard et al. [1] introduced the Food-101 dataset containing over 100,000 images categorized into 101 food categories (e.g. apple pie, paella, risotto) collected from the web. The paper reports overall accuracy of about 56% on the 101-way classification problem, although it varies significantly based on class (e.g. 95% for edamame, 10% for apple pie).

Other researchers have introduced food datasets and techniques that target different applications and challenges. The Pittsburgh Fast Food Image Dataset [5] includes about 4,500 images of 101 foods from 11 fast food restaurants. FoodAI compares food versus non-food images [22]. ChineseFoodNet [6] targets Chinese food items, while UEC-100 focuses on foods from Japan [17]. Kawano et al. [15] study cross-domain food recognition, using images of one type of food to help train classifiers for another. Most of these papers use training and test images collected from the web, which can be highly biased towards idealized photos that people want to share with others. In contrast, Mezgec and Seljak [18] collected real-world image data from Parkinson's disease patients, and obtained about 55% accuracy on a 115-way food classification task.

Identified foods can be further analyzed to estimate food volume, and by extension, the nutrient content of foods. Most approaches for volume estimation include calibration for scale, volume modeling, and referencing against databases [24]. Calibrating for scale is surprisingly difficult due to the scale ambiguity problem in computer vision [11]: it is impossible, from a single two-dimensional image, to estimate both the distance to an object in the three-dimensional scene and the size of the 3D object. To overcome this problem, scale calibration can be approximated using physical fiducial markers such as standardized plates of known diameters [27] or foods of standard size (such as japonica rice grains [10]). In terms of volume mapping, Chae et al. utilized the projection of a known geometric shape over a food item (such as cylindrical shape for glasses of beverages) with 11% mean error [3].

Finally, translating from recognized foods and food volumes to meaningful nutrition information (e.g., calories) depends on the accuracy of available databases that are either maintained by public entities (e.g., the U.N. Food and Agriculture Organization) or private repositories [4]. The Im2Calories system [20] is an example of an attempt to estimate calories from food photos. They considered several subtasks, including segmenting a plate of food into different food items (e.g. eggs, bacon), identifying each item, estimating the food volume, and then computing the total number of calories. Although Im2Calories reported that their CNN volume predictor is accurate for most of the meals, they also reported that they were unable to conduct end-to-end quantitative tests of calorie estimation due to discrepancies in food databases.

3 HEALTH EXPERT REVIEWS ON PHOTO-BASED FOOD DIARY

Researchers in HCI and health informatics have examined the use of photo-based food diaries because they reduce the burden of text-based diaries and provide social support in the pursuit of healthy eating goals when shared on social media [7, 9]. Research has also shown that photo diaries are more reproducible than text-based diaries [12]. From a health expert's point of view, photos provided visual examples to help diabetes educators communicate with patients [16]. The contextual information that photos capture also was found to support IBS patients and people with healthy eating goals to work with health experts to identify triggers or behavior change opportunities [8].

Although the use of photo-based diaries is promising, it is not well understood how computer vision-based systems can support health experts in analyzing photo-based food diaries. We conducted a preliminary study in which 18 dietitians were assigned to review 7-day photo diaries collected by people taking part in a human subjects study. In general, we observed that dietitians looked for eating patterns across meals or days, consistent with what health experts did when using Foodprint in dietary assessments with clients [8]. Dietitians in our study compared the types of food that clients ate in meals versus snacks, at different times during the day, and during different days of the week. They also used color distribution (e.g., green for vegetables versus beige for potatoes) and relative portions (e.g., how many vegetables versus how many proteins clients ate in a day) to determine food variety and balance. Besides food content, dietitians also inferred contextual information presented in the photo such as eating locations, companions, and routines. While some dietitians were interested in clients' overall energy consumption across a day, the focus on caloric limit was minimal.

These findings suggest a significant discrepancy in the problems currently addressed in the computer vision research community (e.g., identifying specific predefined foods, estimating calories, etc.) and what expert dietitians actually look for in food diaries. In contrast to how current computer-vision systems analyze food photos, health experts often look beyond single photo analysis to focus on long-term patterns. They also look beyond the plates to make sense of contextual information during dietary consultations. These discrepancies in approaches and goals suggest several opportunities for future research.

4 CHALLENGES AND OPPORTUNITIES

Current work in computer vision-based food recognition shows promise, but the types of problems it aims to solve may not be widely useful in practice. For example, estimating volumes from food photos is relatively difficult because of the lack of depth information in 2D photographs [14]. This challenge is not unique to computer vision algorithms. Studies show that trained dietetic interns only correctly estimated portion sizes for 30% of food images [13], while untrained individuals have even more difficulty [25]. Computer vision technologies have the potential to solve some recognition problems, but they may also be fundamentally constrained by the limited information present in food images. For example, any analysis of food images, whether by humans or machines, will have difficulty recognizing occluded objects like ingredients inside a sandwich or salad. Despite these challenges, there are ample opportunities for computer vision-based food recognition systems to support individuals and health experts to better use food images to improve health and wellness. Building on current computer vision-based food recognition work, we propose several future directions to better support real-world use.

4.1 Inclusion and Diversity of Food Training Data

Traditional food database-based food diaries often do not include the diverse types of food that individuals consume [9]. In our review and the preliminary study, we found that this is also the case with existing photo image datasets. For example, in a preliminary investigation of photos from an IRB approved study of 80 participants tracking their diet with photos, we found that nearly half contained foods that did not neatly fall into the 101 categories of the popular Food-101 dataset [1].

While not all datasets are limited in the same way, system designers and developers need to consider the diversity of food that people have access to and choose to eat. The low presence of particular types of food in a training dataset can result in low recognition rates. When these systems are adopted in dietary assessment, the inaccuracies might lead to incorrect diagnoses or inappropriate recommendations. These errors may not be uniformly distributed across the population, but instead affect people of specific backgrounds or socioeconomic groups depending on the foods they eat. More research should strive for ways to curate and adopt more diverse datasets. Research should also recognize the limitations that current datasets inherit and consider them in the overall algorithm and system design.

4.2 The Social-Technical Gap of Food Image Recognition

Much research has focused on building food image recognition techniques and improving their accuracy. However, there is a gap between computer vision research and the types of problems this research is meant to address in real-world scenarios. For example, many existing datasets only include restaurant foods and professional photos, while in real life, people often prepare their own food at home and take photos in a variety of ways. As shown in previous research in database-based food diaries [9], the low recognition rate of everyday foods could even potentially discourage people from eating foods aligned with their health goals (e.g. homemade food), leading them instead toward foods that are easily recognizable by automatic diaries to reduce burden (e.g., restaurant food or package food).

Similarly, most computer vision work focuses on recognizing food content from single photos. In real life, many health goals and conditions rely on long-term eating behavior change or management. Recognition based on single instances of eating may risk missing the overall picture of individual behavior and patterns. We see an opportunity for food recognition research to better understand longitudinal eating patterns, contexts, and behaviors beyond a single plate, to support more individualized assessment and recommendations. This longer-term approach may actually *ease* the automated recognition challenges because the system can use evidence from multiple photos to resolve visual ambiguities and uncertainties (e.g. by customizing its model, over time, to each individual and the foods they tend to eat).

4.3 Human-AI Collaboration in Dietary Assessment

Leveraging computer vision could have many benefits, especially for people and health providers in low-resource communities. These systems can also provide just-in-time support when providers are not available. However, many of the health goals and concerns that computer vision-based dietary assessment can be applied to require complex considerations beyond single food photo recognition, such as individual preferences and constraints that influence whether and how they adopt everyday behavior change or management strategies. For example, people with eating disorders may require both dietary and psychological consultation [21]. Simply replacing experts with recommendations based on food image recognition, even if done accurately, may risk overlooking important factors supporting health management. A better approach might be to design computer vision-based dietary assessment systems to support dietitians and nutrition experts working with individuals. Promoting collaborations between human experts and systems may decrease the manual assessment effort and time, allowing experts to spend more time interacting with individuals. These collaborations, however, require a better understanding of the support that experts need in dietary assessment and how they work with individuals.

5 CONCLUSION

While computer vision algorithms have greatly advanced in recent years, there are still challenges in adopting these systems in real-world use. In this position paper, we proposed three research directions in supporting computer vision-based dietary assessment. First, we need to recognize the bias created by the training data in creating recognition models and their potential influence on dietary assessment. Second, dietary management requires more than an accurate estimation of nutrients, portions, and calories. We need to understand the problems and needs of individuals and think about how we can apply these technologies in supporting these needs. Finally, we need to examine a more holistic approach to support individual health goals, by understanding how computer vision algorithms can collaborate and complement human experts, instead of trying to replace them.

6 ACKNOWLEDGMENTS

This work was supported in part by the Precision Health Initiative at Indiana University, and by an National Science Foundation Research Experiences for Undergraduates (REU) program (IIS-1852294).

REFERENCES

[1] Lukas Bossard, Matthieu Guillaumin, and Luc Van Gool. 2014. Food-101 – Mining Discriminative Components with Random Forests. In *European Conference on Computer Vision*.

[2] Vieira Bruno and Cui Juan Silva Resende. 2017. A survey on automated food monitoring and dietary management systems. *Journal of health & medical informatics* 8, 3 (2017).

[3] J. Chae, I. Woo, S. Kim, R. Maciejewski, F. Zhu, E. J. Delp, C. J. Boushey, and D. S. Ebert. 2011. Volume Estimation Using Food Specific Shape Templates in Mobile Image-Based Dietary Assessment. *Proc SPIE Int Soc Opt Eng* 7873 (Feb 2011), 78730K.

[4] U.R. Charrondiere, D. Haytowitz, and B. Stadlmayr. 2012. FAO/INFOODS Density Database Version 2.0. *Food and Agriculture Organization of the United Nations Technical Workshop Report 2012* (2012).

[5] Mei Chen, Kapil Dhinga, Wen Wu, Lei Yang, Rahul Sukthankar, and Jie Yang. 2009. PFID: Pittsburgh fast-food image dataset. In *IEEE Conference on Computer Vision and Pattern Recognition*.

[6] Xin Chen, Hua Zhou, Yu Zhu, and Liang Diao. 2017. ChineseFoodNet: A large-scale Image Dataset for Chinese Food Recognition. *arXiv preprint arXiv:1705.02743* (2017).

[7] Chia-Fang Chung, Elena Agapie, Jessica Schroeder, Sonali Mishra, James Fogarty, and Sean A Munson. 2017. When personal tracking becomes social: Examining the use of Instagram for healthy eating. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*. 1674–1687.

[8] Chia-Fang Chung, Qiaosi Wang, Jessica Schroeder, Allison Cole, Jasmine Zia, James Fogarty, and Sean A Munson. 2019. Identifying and planning for individualized change: Patient-provider collaboration using lightweight food diaries in healthy eating and irritable bowel syndrome. *Proceedings of the ACM on interactive, mobile, wearable and ubiquitous technologies* 3, 1 (2019), 1–27.

[9] Felicia Cordeiro, Daniel A Epstein, Edison Thomaz, Elizabeth Bales, Arvind K Jagannathan, Gregory D Abowd, and James Fogarty. 2015. Barriers and negative nudges: Exploring challenges in food journaling. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*. ACM, 1159–1162.

[10] Takumi Ege, Wataru Shimoda, and Keiji Yanai. 2019. A New Large-Scale Food Image Segmentation Dataset and Its Application to Food Calorie Estimation Based on Grains of Rice. In *Proceedings of the 5th International Workshop on Multimedia Assisted Dietary Management* (Nice, France) *(MADiMa '19)*. Association for Computing Machinery, New York, NY, USA, 82âĂŞ87. https://doi.org/10.1145/3347448.3357162

[11] Isaac Esteban, Leo Dorst, and Judith Dijk. 2010. Closed form solution for the scale ambiguity problem in monocular visual odometry. In *International Conference on Intelligent Robotics and Applications*. Springer, 665–679.

[12] Juan M Fontana, Zhaoxing Pan, Edward S Sazonov, Megan A McCrory, J Graham Thomas, Kelli S McGrane, Tyson Marden, and Janine A Higgins. 2020. Reproducibility of dietary intake measurement from diet diaries, photographic food records, and a novel sensor method. *Frontiers in Nutrition* 7 (2020).

[13] Erica Howes, Carol J Boushey, Deborah A Kerr, Emily J Tomayko, and Mary Cluskey. 2017. Image-based dietary assessment ability of dietetics students and interns. *Nutrients* 9, 2 (2017), 114.

[14] W. Jia, Y. Yue, J. D. Fernstrom, Z. Zhang, Y. Yang, and M. Sun. 2012. 3D localization of circular feature in 2D image and application to food volume estimation. In *2012 Annual International Conference of the IEEE Engineering in Medicine and Biology Society*. 4545–4548. https://doi.org/10.1109/EMBC.2012.6346978

[15] Yoshiyuki Kawano and Keiji Yanai. 2014. Automatic Expansion of a Food Image Dataset Leveraging Existing Categories with Domain Adaptation. In *European Conference on Computer Vision*.

[16] Lena Mamykina, Elizabeth Mynatt, Patricia Davidson, and Daniel Greenblatt. 2008. MAHI: investigation of social scaffolding for reflective thinking in diabetes management. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. 477–486.

[17] Y. Matsuda and K. Yanai. 2012. Multiple-food recognition considering co-occurrence employing manifold ranking. In *IAPR International Conference on Pattern Recognition*.

[18] Simon Mezgec and Barbara Korousic Seljak. 2017. NutriNet: A Deep Learning Food and Drink Image Recognition System for Dietary Assessment. *Nutrients* 9 (2017), 657. Issue 7.

[19] Weiqing Min, Shuqiang Jiang, Linhu Liu, Yong Rui, and Ramesh Jain. 2019. A Survey on Food Computing. *Comput. Surveys* 52, 92 (2019).

[20] Austin Myers, Nick Johnston, Vivek Rathod, Anoop Korattikara, Alex Gorban, Nathan Silberman, Sergio Guadarrama, George Papandreou, Jonathan Huang, and Kevin Murphy. 2015. Im2Calories: Towards an automated mobile vision food diary. In *Proceedings of the IEEE International Conference on Computer Vision*.

[21] Nicola Rance, Naomi P Moller, and Victoria Clarke. 2017. âĂŸEating disorders are not about food, theyâĂŹre about lifeâĂŹ: Client perspectives on anorexia nervosa treatment. *Journal of Health Psychology* 22, 5 (2017), 582–594. https://doi.org/10.1177/1359105315609088 arXiv:https://doi.org/10.1177/1359105315609088 PMID: 26446375.

[22] Doyen Sahoo, Wang Hao, Shu Ke, Wu Xiongwei, Hung Le, Palakorn Achananuparp, Ee-Peng Lim, and Steven C. H. Hoi. 2019. *FoodAI: Food Image Recognition via Deep Learning for Smart Food Logging*. Association for Computing Machinery, New York, NY, USA, 2260âĂŞ2268. https://doi.org/10.1145/3292500.3330734

[23] Neha Sharma, Vibhor Jain, and Anju Mishra. 2018. An Analysis Of Convolutional Neural Networks For Image Classification. *Procedia Computer Science* 132 (2018), 377–384. https://doi.org/10.1016/j.procs.2018.05.198 International Conference on Computational Intelligence and Data Science.

[24] Wesley Tay, Bhupinder Kaur, Rina Quek, Joseph Lim, and Christiani Jeyakumar Henry. 2020. Current Developments in Digital Quantitative Volume Estimation for the Optimisation of Dietary Assessment. *Nutrients* 12, 4 (2020). https://doi.org/10.3390/nu12041167

[25] Frances E Thompson and Amy F Subar. 2017. Dietary assessment methodology. In *Nutrition in the Prevention and Treatment of Disease*. Elsevier, 5–48.

[26] Kang Tong, Yiquan Wu, and Fei Zhou. 2020. Recent advances in small object detection based on deep learning: A review. *Image and Vision Computing* 97 (2020), 103910. https://doi.org/10.1016/j.imavis.2020.103910

[27] Y. Yue, W. Jia, and M. Sun. 2012. Measurement of food volume based on single 2-D image without conventional camera calibration. In *2012 Annual International Conference of the IEEE Engineering in Medicine and Biology Society*. 2166–2169. https://doi.org/10.1109/EMBC.2012.6346390