Evaluating Performance and Robustness of AI-based Calories and Nutrients Tracking from Images

Franz Rott
University of Ulm
Ulm, Germany
franz.rott@uni-ulm.de

ABSTRACT

1 INTRODUCTION

According to an estimation by the WHO 16% of the global adult population was obese in 2022, a proportion which more than doubled since 1990 [14] and is likely to increase further [6]. Obesity is accompanied by an elevated risk for cardiovascular diseases, type 2 diabetes, and heightened mortality [14]. Especially during childhood it also harms psychosocial well-being and increases the likelihood of being bullied and stigmatized [14]. Besides grave personal impairments, obesity also has a negative impact on society. The yearly global costs of overweight and obesity are estimated to be US$ 3 trillion by 2030 and over US$ 18 trillion by 2060. Obesity results from an imbalance between energy intake and expenditure and is influenced by both psycho-social and genetic factors [14]. As a study by Zhou et al. [22] revealed, humans are inherently bad at estimating the calories contained in a meal. On average only for 25% of dishes, the estimation error was less than 20%. A possible solution is to replace the error-prone human estimation with an AI model. While most current approaches need the user to upload photos of their meals [15–17] first implementations use online footage from hat or body-worn cameras [2]

In both cases, a backend AI is used to estimate calories or nutrients. Research is at an early stage, and little is known about the robustness of these approaches. We address this research gap by assessing both the model's performance and the influencing photographic factors.

2 RELATED WORK

2.1 State of the Art AI Models for Calorie Detection

In 2009 the Pittsburgh Fast-Food Image Dataset (PFID) was introduced containing 4,545 static images, 606 stereo image pairs, 303 360° videos, and 27 privacy-conscious videos capturing eating events [4]. The dataset consists of 101 foods from 11 popular fast food chains with 3 instances per class and 8 images per instance [4]. For half a decade it remained the only publicly available dataset for food classification. AI models trained on this data include Support Vector Machine (SVM) classifiers by Chen et al. [4] and Yang et al. [20] which depending on whether color histogram methods were used achieved a modest accuracy of 11% and 28%. Although marking an important step forward for automatic calorie detection, the dataset was limited in its scope (only containing fast food images) and created under artificial and laboratory conditions.

Both issues were addressed by Bossard et al. [3]. They created the Food-101 dataset containing 1,000 photos of each of the 101 most popular and consistently named dishes. It was used to train multiple AI models for image classification and calorie detection. Liu et al. [13] tested various deep learning architectures on food recognition, with the best performing being a CNN, reaching a top-1 accuracy of 77.4%. Ciocca et al. [5] chose a similar approach comparing different CNN architectures. Residual Networks (ResNet-50) performed best with a top-1 accuracy of 82.5%. VijayaKumari et al. [19] used Efficientnetb0, a transfer learning technique, to reach a categorization accuracy of 80%.

Other studies follow similar approaches partly with self-generated and not publically available datasets. Joutou and Yanai [11] used Multi-Kernel-Learning (MKL) to achieve a classification rate of 61.3% for 50 kinds of foods. Hoashi et al. [10] created the Food85 dataset and achieved a food classification accuracy of 62.5% by combining MKL and Bag-of-Features (BoF) Other food category recognition systems can be found in the review by Zhang et al. [21].

Although accurate food classification is important, some use cases including dietary assessment require further information like calorie and nutrient content.Haris et al. [8] utilized deep learning techniques, especially CNN combined with ResNet50 for feature extraction, to analyze images, and identify food items with an accuracy of 75.8%. While the approach can also estimate portion sizes and predict the caloric content, no metrics were reported.

By now there exist similar open-source food datasets like the Open Images V6-Food Dataset [1], School Lunch Dataset, Vietnamese Food Dataset [2], or MAFood-121 Dataset [1]. Together with the Food-101 Dataset [3] they were used by Han et al. [7] to train YOLOv8, a food items detection model with an accuracy of 75.4%. Again the approach is generally capable of calorie prediction but without reporting performance. Liang and Li [12] propose a calorie estimation method that leverages Faster R-CNN for food. The system requires two images per food item (top and side views). It uses volume estimation combined with density and energy values to calculate calories with an estimation error generally below ±20%. Thames et al. [18] trained a computer vision algorithm for predicting calories and macronutrients of real-world dishes with an accuracy exceeding those of professional nutritionists.

To summarize there is an emerging field of research regarding image-based AI calorie detection. The main scientific focus lies on food classification. Currently, there is little evidence of how well the improving classification also translates to accurate calorie prediction. Studies that address this research question neglect to explore how photographic variables influence performance. We aim to close this research gap by experimenting with possible deciding factors like the angle from which an image is taken and image characteristics like lighting and contrast. Both factors are speculated to have an impact because of the presumed imbalance in the dataset. For our study, we focus on the state-of-the-art AI model trained by [8], because of its accessibility via its API. The

[1] https://www.kaggle.com/datasets/programmerrdai/open-images-v6
[2] https://universe.roboflow.com/nhh/vietnamese-food/dataset/1

model and the Food-101 dataset it was trained on are described in more detail in the following Sections.

2.2 Food-101 Dataset

The Food-101 Dataset was created by Bossard et al. [3] to make a publically available food recognition dataset that is less artificial, contains noise is generally less laboratory. This was implemented by downloading randomly chosen images of the 101 most popular and consistently named dishes from foodspotting.com. Each class consists of 750 train and 250 manually cleaned test images with a maximum side length of 512 pixels. The noise, especially false labeling and intense coloring was kept for the training data set to account for some real-world noise.

Although marking a clear improvement over former public datasets like the PFID Chen et al. [4], Food-101 still has some potential shortcomings. Since coming from a community-driven dish discovery website one must assume that most meals are photographed approximately from above and with "normal" settings to appeal to the audience. This might not properly represent the images used for calorie estimation, especially if they are captured by a body or hat-worn camera. Furthermore, 101 dishes represent only a fraction of the actual food variety.

2.3 Foodvisor AI and API

Besides others the Food-101 Dataset was used by Haris et al. [8] to train and test the Foodvisor AI model. To augment the training data random cropping and horizontal flipping were used. The trained CNN model was paired with ResNet50 model He et al. [9] for feature extraction. Introduced in 2016, ResNet50 enhances the performance by incorporating residual mapping and skip connections, optimizing the model's depth, accuracy, and robustness. This combined approach achieved a classification accuracy of 75.8%. While no metrics on portion size and calorie estimation were reported, these are by now further capabilities of the Foodvisor AI which is accessible via their chargeable API.

As a state-of-the-art AI model for food classification, we are aiming to test the calorie estimation capabilities of this model and its robustness.

3 EXPERIMENTAL APPARATUS

3.1 Dataset

For testing the capabilities of the Foodvisor-AI, we base our dataset on the Nutrition5k dataset Thames et al. [18]. It consists of video streams, depth images, component weights, and nutritional content labels of 5,000 diverse real-world food dishes. For constructing our test dataset we extract the image taken from straight above and from a 45° angle for 100 randomly chosen dishes. For each of the 200 images, we create 7 variations using the ... Python library: normal, +10% brightness, -10% brightness, +10% contrast, -10% contrast, +10% saturation, -10% saturation. We focus on these exact image settings since they vary the most in everyday settings since they are directly influenced by changing light conditions, like time of day or indoor versus outdoor environments. Our final dataset therefore consists of 1,400 images. We restrict it to that amount to keep the API costs reasonable. We go more into detail about how we utilize the API in the following Section.

3.2 Procedure

We build a connection to the Foodvisor AI via its API to individually prompt the images of our created dataset. We retrieve the amount of estimated calories and macronutrients and save the result for later analysis. Because of the two orthogonal dimensions *angle* with two values and *image settings* with seven, we have a total of 14 groups we can compare. We are especially interested in the accuracy within each group and the variability between groups. We aim to identify groups for which the prediction is rather inaccurate to drive the improvement of AI training (data) and derive actionable insights for real-world (life) usage of these models.

3.3 Measures

To assess performance, we use the mean absolute error (MAE)

$$\mathrm{MAE} = \frac{1}{n}\sum_{i=1}^{n} |y_i - \hat{y}_i|$$

and mean absolute percentage error (MAPE)

$$\mathrm{MAPE} = \frac{1}{n}\sum_{i=1}^{n} \left|\frac{y_i - \hat{y}_i}{y_i}\right| \times 100\%$$

for predicting calories and nutrients. Comparing these metrics across different angles and image settings allows us to also draw conclusions about the robustness.

4 RESULTS

5 DISCUSSION

5.1 Key Scientific Insights

5.2 Threat to Validity and Limitations

5.3 Future Work and Impact

6 CONCLUSION

REFERENCES

[1] Eduardo Aguilar, Marc Bolaños, and Petia Radeva. 2019. Regularized uncertainty-based multi-task learning model for food analysis. *Journal of Visual Communication and Image Representation* 60 (2019), 360–370.

[2] Shengjie Bi and David Kotz. 2022. Eating detection with a head-mounted video camera. In *2022 IEEE 10th International Conference on Healthcare Informatics (ICHI)*. IEEE, 60–66.

[3] Lukas Bossard, Matthieu Guillaumin, and Luc Van Gool. 2014. Food-101–mining discriminative components with random forests. In *Computer vision–ECCV 2014: 13th European conference, zurich, Switzerland, September 6-12, 2014, proceedings, part VI 13*. Springer, 446–461.

[4] Mei Chen, Kapil Dhingra, Wen Wu, Lei Yang, Rahul Sukthankar, and Jie Yang. 2009. PFID: Pittsburgh fast-food image dataset. In *2009 16th IEEE International Conference on Image Processing (ICIP)*. IEEE, 289–292.

[5] Gianluigi Ciocca, Paolo Napoletano, and Raimondo Schettini. 2018. CNN-based features for retrieval and classification of food images. *Computer Vision and Image Understanding* 176 (2018), 70–77.

[6] Richard Dobbs, Corinne Sawers, Fraser Thompson, James Manyika, Jonathan R Woetzel, Peter Child, Sorcha McKenna, and Angela Spatharou. 2014. *Overcoming obesity: an initial economic analysis*. McKinsey global institute.

[7] Michelle Han, Junyao Chen, and Zhengyuan Zhou. 2024. NutrifyAI: An AI-Powered System for Real-Time Food Detection, Nutritional Analysis, and Personalized Meal Recommendations. arXiv:2408.10532 [cs.CV] https://arxiv.org/abs/2408.10532

[8] Mohammad Haris, Mohd Sarim, Vaibhav Mishra, and Garima Singh. 2024. FoodVisor: A Food Calorie Estimation System. In *2024 Second International Conference on Advances in Information Technology (ICAIT)*, Vol. 1. IEEE, 1–7.

[9] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. 2016. Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*. 770–778.

[10] Hajime Hoashi, Taichi Joutou, and Keiji Yanai. 2010. Image recognition of 85 food categories by feature fusion. In *2010 IEEE International Symposium on Multimedia*. IEEE, 296–301.

[11] Taichi Joutou and Keiji Yanai. 2009. A food image recognition system with multiple kernel learning. In *2009 16th IEEE International Conference on Image Processing (ICIP)*. IEEE, 285–288.

[12] Yanchao Liang and Jianhua Li. 2017. Computer vision-based food calorie estimation: dataset, method, and experiment. *arXiv preprint arXiv:1705.07632* (2017).

[13] Chang Liu, Yu Cao, Yan Luo, Guanling Chen, Vinod Vokkarane, and Yunsheng Ma. 2016. Deepfood: Deep learning-based food image recognition for computer-aided dietary assessment. In *Inclusive Smart Cities and Digital Health: 14th International Conference on Smart Homes and Health Telematics, ICOST 2016, Wuhan, China, May 25-27, 2016. Proceedings 14*. Springer, 37–48.

[14] World Health Organization et al. 2024. Obesity and overweight. 2024: https://www. who. int/news-room/fact-sheets/detail/obesity-and-overweight. *Updated January* (2024).

[15] Ioannis Papathanail, Jana Brühlmann, Maria F Vasiloglou, Thomai Stathopoulou, Aristomenis K Exadaktylos, Zeno Stanga, Thomas Münzer, and Stavroula Mougiakakou. 2021. Evaluation of a novel artificial intelligence system to monitor and assess energy and macronutrient intake in hospitalised older patients. *Nutrients* 13, 12 (2021), 4539.

[16] G Ramkumar, CB Venkatramanan, V Manonmani, S Palanivel, Ata Kishore Kumar, et al. 2023. A Real-time Food Image Recognition System to Predict the Calories by Using Intelligent Deep Learning Strategy. In *2023 International Conference on Research Methodologies in Knowledge Management, Artificial Intelligence and Telecommunication Engineering (RMKMATE)*. IEEE, 1–7.

[17] S Sapna, KM Yaswanth, Praveen Kumar, et al. 2023. Calorie Estimation of Food and Beverages using Deep Learning. In *2023 7th International Conference on Computing Methodologies and Communication (ICCMC)*. IEEE, 324–329.

[18] Quin Thames, Arjun Karpur, Wade Norris, Fangting Xia, Liviu Panait, Tobias Weyand, and Jack Sim. 2021. Nutrition5k: Towards automatic nutritional understanding of generic food. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*. 8903–8911.

[19] G VijayaKumari, Priyanka Vutkur, and P Vishwanath. 2022. Food classification using transfer learning technique. *Global transitions proceedings* 3, 1 (2022), 225–229.

[20] Shulin Yang, Mei Chen, Dean Pomerleau, and Rahul Sukthankar. 2010. Food recognition using statistics of pairwise local features. In *2010 IEEE computer society conference on computer vision and pattern recognition*. IEEE, 2249–2256.

[21] Yudong Zhang, Lijia Deng, Hengde Zhu, Wei Wang, Zeyu Ren, Qinghua Zhou, Siyuan Lu, Shiting Sun, Ziquan Zhu, Juan Manuel Gorriz, et al. 2023. Deep learning in food category recognition. *Information Fusion* 98 (2023), 101859.

[22] Jun Zhou, Dane Bell, Sabrina Nusrat, Melanie Hingle, Mihai Surdeanu, Stephen Kobourov, et al. 2018. Calorie estimation from pictures of food: Crowdsourcing study. *Interactive journal of medical research* 7, 2 (2018), e9359.