STM JOURNALS

Journal of
Open Source Developments

ISSN: 2395-6704
Volume 11, Issue 2, 2024
May–August
DOI (Journal): 10.37591/JoOSD

https://journals.stmjournals.com/joosd

Review JoOSD

Depiction-inspired Recipe Generator Using Deep Learning

Om Sanjay Awari[1,*], Shivaraj Mahendra Ghangale[1], Janmejay Aklesh Maurya[1], Aishwarya Ajit Ghangale[1], Keerti Kharatmol[2]

Abstract

Machine learning has become a crucial part of modern life, influencing various domains. Its applications range from enhancing data-driven business decisions to enabling autonomous vehicles. Advances in machine learning have brought about notable changes in how we interact with technology. In the culinary world, the idea of creating food recipes from images has gained increasing interest. This entails the development of innovative systems that seamlessly convert visual input, such as images of dishes, into comprehensive culinary instructions. This process involves a fusion of advanced computer vision techniques and natural language generation algorithms, enabling a bridge between visual data and textual information. Several cutting-edge systems have been proposed to tackle the challenge of generating recipes from images. These systems employ sophisticated methodologies that leverage object detection and recognition to accurately identify ingredients and quantities within the images. By consulting extensive recipe databases, these systems generate step-by-step cooking instructions. The applications of such systems are far-reaching. They serve a wide range of users, from seasoned chefs looking for new ideas to beginners needing cooking advice. This technology introduces a novel aspect to cooking by making culinary knowledge more accessible, encouraging creativity, and improving the cooking experience. The advantages of using machine learning for recipe generation are numerous. It acts as a digital sous-chef, offering personalized recommendations tailored to user preferences and dietary needs. Additionally, the integration of user feedback and continuous refinement through advanced machine learning techniques ensures the adaptability and accuracy of these systems. Moreover, the seamless integration of such systems with smart cooking devices promises to revolutionize the way individuals approach cooking, making it more accessible, interactive, and enjoyable for everyone.

Keywords: Deep learning, recipe generation, cooking instructions, food images, culinary domain

***Author for Correspondence**
Om Sanjay Awari
E-mail: omawari@kccemsr.edu.in

[1]Student, Department of Computer Science and Engineering, K.C. College of Engineering & Management Studies & Research, Thane, Maharashtra, India
[2]Assistant Professor, Department of Computer Science and Engineering, K.C. College of Engineering & Management Studies & Research, Thane, Maharashtra, India

Received Date: May 06, 2024
Accepted Date: July 31, 2024
Published Date: August 07, 2024

Citation: Om Sanjay Awari, Shivaraj Mahendra Ghangale, Janmejay Aklesh Maurya, Aishwarya Ajit Ghangale, Keerti Kharatmol. Depiction-inspired Recipe Generator Using Deep Learning. Journal of Open Source Developments. 2024; 11(2): 47–55p.

INTRODUCTION

In the sector of artificial intelligence, the convergence of computer vision and natural language processing has paved the way for innovative solutions across diverse domains [1]. One such entertaining application emerges in the culinary domain, where the fusion of image classification, recipe extraction, and multi-language translation is ready to redefine the landscape of automated food recipe generation. This paper introduces a comprehensive model designed to holistically approach the culinary creative process by first classifying food images, extracting detailed recipes and ingredients from an extensive dataset, and subsequently translating the generated recipes into multiple languages. The initial phase of our model involves leveraging state-of-the-art image classification techniques to accurately identify and categorize diverse food items. This classification serves as the foundational step towards understanding the visual context of a dish, facilitating subsequent recipe extraction. Our model is trained on a meticulously curated dataset encompassing a vast array of cuisines, ensuring robust performance across various culinary traditions and regional specialties [2].

Following image classification, our model employs advanced natural language processing techniques to extract intricate details of the recipe and its corresponding ingredients. This phase involves the synthesis of information from multiple sources, encompassing cooking blogs, recipe databases, and culinary literature, to enhance the model's understanding of the relationships between ingredients and their proportions. The result is a comprehensive and coherent representation of the culinary process, poised to inspire creativity in the kitchen. Moreover, our model extends its capabilities beyond the boundaries of language by incorporating a multilingual translation component. Recognizing the global diversity of culinary preferences, the model seamlessly translates the generated recipes into multiple languages, ensuring accessibility and inclusivity [3]. This groundbreaking feature not only broadens the reach of culinary inspiration but also facilitates cross-cultural exchange and appreciation of diverse gastronomic traditions.

LITERATURE REVIEW

Tracking food intake has long been an open problem in nutritional science. Manual tracking is known to be bulky and imperfect as reported in reference [4]. A viable solution adopted by commercial apps is by automatic identification of food content from picture and derivation of nutrition facts as well as calories by matching to food composition table. Current technologies, nevertheless, are limited to barcode identification of raw ingredients and visual recognition to a limited number of standardized cooked cafeteria foods from chained restaurants [5]. The literature survey emphasizes the importance of incorporating rich food attributes, such as ingredients, cooking methods, and flavors, to enhance the accuracy and relevance of cross-modal recipe retrieval systems.

Traditionally, the image-to-recipe problem has been approached as a retrieval task, where a recipe is fetched from a fixed dataset based on how similar the image is to those in the embedding space. The effectiveness of such systems relies heavily on the size and diversity of the dataset, as well as the quality of the learned embeddings. To address the limitations of retrieval-based systems, an alternative approach is to treat the image-to-recipe problem as a conditional generation task. In this paper, we propose a system that generates a complete cooking recipe, including a title, ingredients, and instructions, from our model [6]. For example, extracting food information like ingredients and recipes from an image of a meal could assist in tracking daily intake and managing dietary habits. Beyond logging food intake, computational food analysis is also vital for understanding the functional similarity of ingredients and predicting meal preferences [7]. We can describe our task as generating a highly realistic ingredients and recipe instructions for the given image. Generating ingredients and recipe instruction from the image has belonged to the category of image-to-text tasks. We concentrate on developing joint representations for both textual and visual modalities, specifically in the context of food images, ingredients, and cooking instructions related to those images. Apart from the significance of this domain to cooking arts, these insights are applied to social media and cultural preservation in term of ethnic food presentations and recommendations. For the development of automatic and analytical methods there is a dire need for diverse and multi perspective datasets [8]. The next section gives an idea of the datasets that we are created to train our model.

Dataset

The image dataset which we created for the training of the model were from the images collected from multiple websites and the multiple resources. Figure 1 shows the images that we collected for the *gulab jamun* dataset to train the model. Similarly, we collect the images and create dataset for multiple sweet dishes [9].

Figure 1. *Gulab jamun* dataset.

METHODOLOGY

Deep learning recipe generation leverages advanced algorithms to create unique and tailored culinary recipes based on vast datasets of ingredients and cooking techniques as shown in Figure 2.

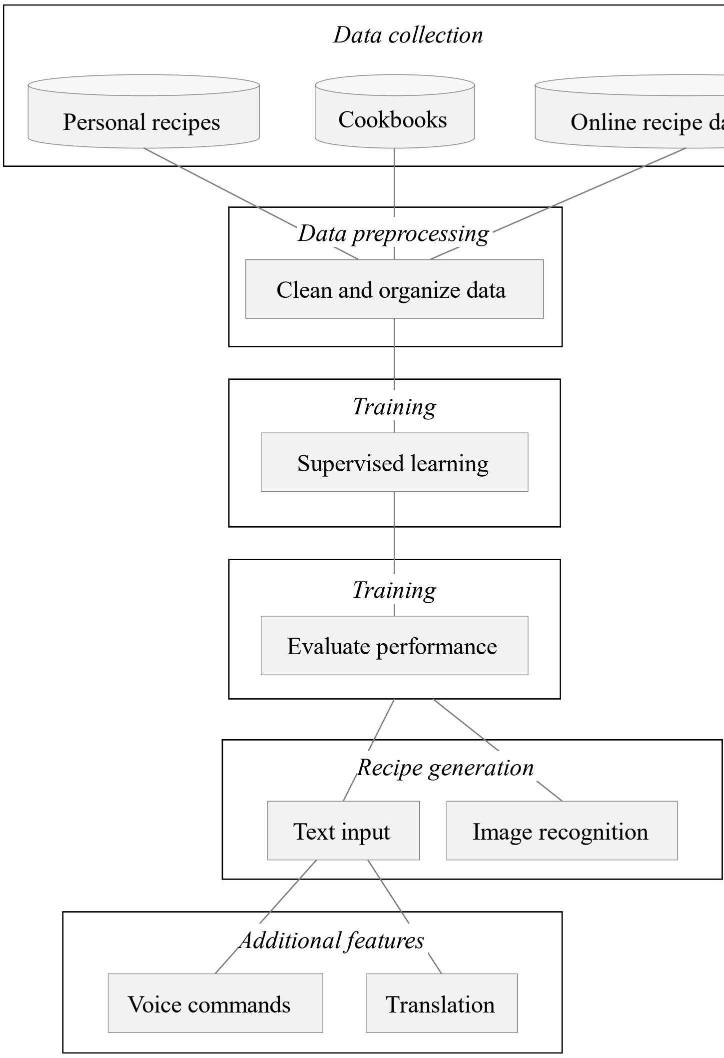

Figure 2. Methodology of deep learning recipe generation.

1. *Data collection:* Here we collected the images of the dishes for multiple sweet dishes. We also collects/created the recipe data for our project. We collect the recipe data from the online recipe databases.
2. *Data preprocessing:* During the data preprocessing step, we processed our image data by collecting more than 20 publicly available images for each dish. Subsequently, we utilized this image dataset for training our model.
3. *Training:* We trained our model using a convolutional neural network (CNN) algorithm, which is tailored for tasks involving image recognition. The images we used for training were created for this purpose [10]. CNNs are a type of artificial neural network known for their ability to recognize patterns in images by leveraging deep learning techniques.
4. *Testing:* During testing, we assess our model's performance by inputting various food images and examining the accuracy of the outputs it provides. We evaluate how well our model performs in recognizing different foods.
5. *Recipe generation:* In the Recipe Generation step, the model creates recipes based on the ingredients it identifies from the provided image. It examines the image, identifies the ingredients from a dataset, and then generates a recipe corresponding to the food image given as input.
6. *Additional features:* In our additional features, we offer three options. Firstly, there is a voice feature where our system can speak out the recipe provided by the model [11]. Secondly, there is a language converter feature that allows users to translate the recipe into their preferred language for better understanding. Lastly, there is a YouTube video link feature that provides a link to a video demonstrating the recipe's preparation.

ALGORITHM (PROCESS)

Now we will see the Algorithm/Process of the model that we created [12].

1. *Step 1:* For the first step, we made our own dataset for the machine to classify images. We did not use an existing online dataset. Also, we created a dataset specifically for the recipes used in our model. We have provided an example of this recipe dataset we created.
2. *Step 2:* If it is your first time using our website, you will need to fill out a signup form before you can log in. After signing up, you can use your credentials to log in at any time and access our services. If you have previously visited our site, simply log in with your existing credentials to access the features.
3. *Step 3:* After logging in, you have two options. You can either classify an image and get a recipe, or if you already have the ingredients, you can get the recipe name and a link to the corresponding YouTube video.
4. *Step 4:* If you pick the classification option, you will be taken to our model page. There, you can upload a picture of food. Our model will then tell you the name of the recipe, what ingredients you need, how to make it, and even give you a link to a YouTube video showing the recipe being made.
5. *Step 5:* If you do not understand English, no worries! You can change the language to one you are comfortable with. And if you prefer, the system can even speak out the recipe for you after you have received it.
6. *Step 6:* And in step 2 if you choose the second option, where you provide the ingredients you have, the model will analyze them and give you recipe names that match those ingredients. It will also include a YouTube link for each recipe, allowing you to watch the preparation process.

RESULTS

First, we have created a home page for our model. In this page we have given the four buttons in the panel. And those buttons are Home, Contact, Sign In, and About. After clicking on the contact button, it will show the Contact us form which contains the fields as Name, Email, and Message, which the user has to give, or the user can ask a question also using this form [13]. Then after clicking on the About button then it will show the information related to our project or model which we are created for the food recipe generation. Clicking the Sign In button will redirect the user to the sign-in page [14].

We have created a page for new users to sign up (Figure 3). This page is created for our model. The page contains a form in that there are fields which get the information from the new users [15]. The sign-up form has fields of username, email, password, and confirm password. After all the fields are filled in, there is a Sign Up button for creating the account of user [16].

We have created a Sign In page (Figure 4) for the users to sign in and then they can go for the main page of our model [17]. This page contains one form which has the fields like username and password as usual any login form has. After filling all the fields, the user needs to click on the Sign In button. After that the user directly goes to the main page for our model [18].

We have created a page for the users to upload an image (Figure 5). After uploading the image, it analyzes the image and gives the ingredients and recipe which is in the dataset we have provided in the code [19].

Figure 6 shows the model providing the ingredients and recipes for the image that the user provided to classify. After classifying and providing the recipes for the image, it provides one YouTube link for that recipe [20].

Figure 3. Sign up page.

Figure 4. Sign In page.

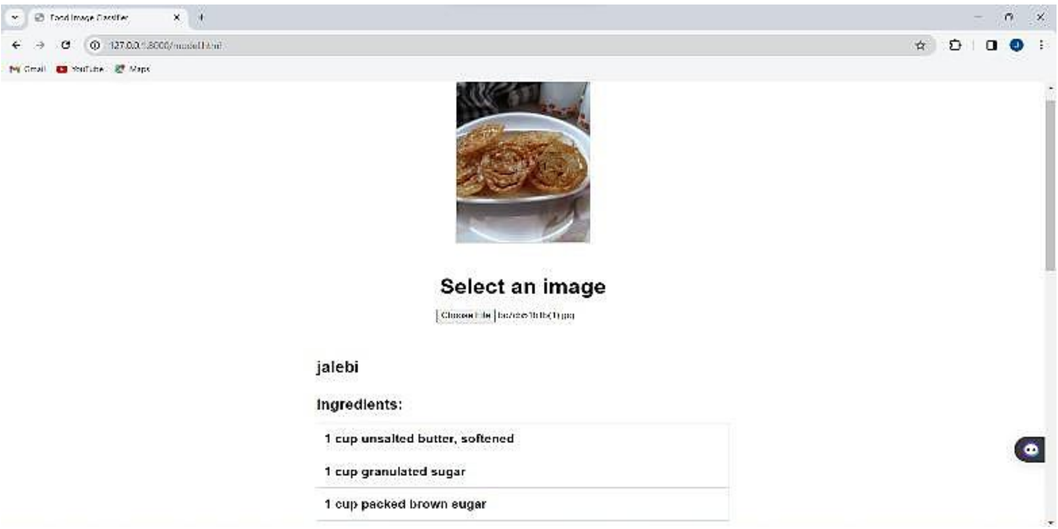

Figure 5. Model page.

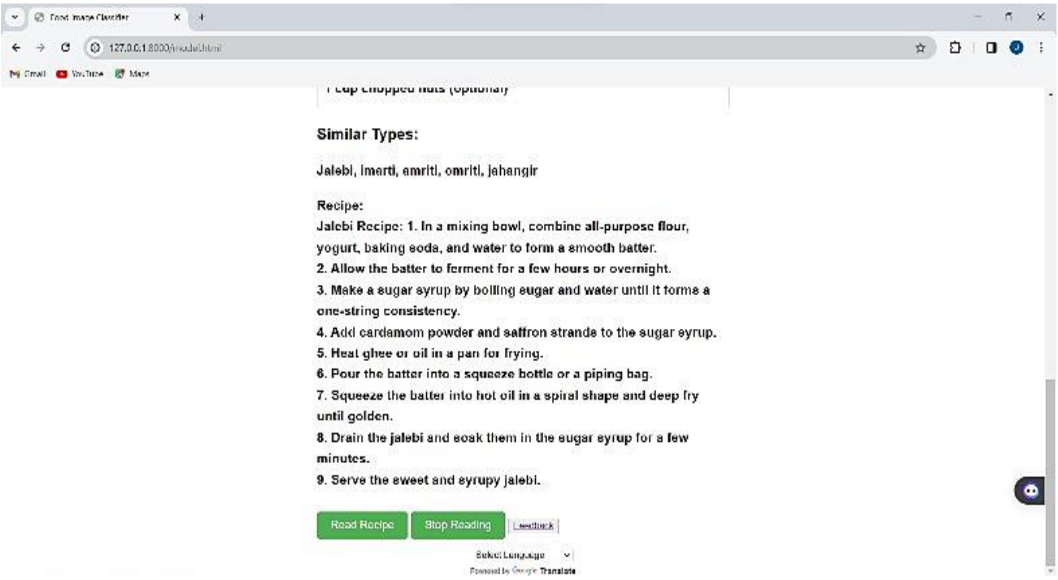

Figure 6. Model page.

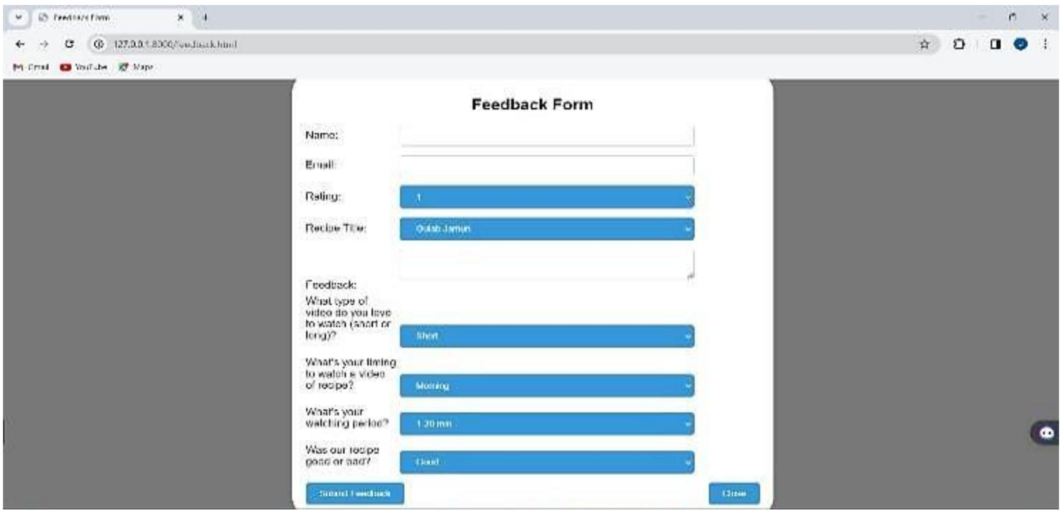

Figure 7. Feedback form page.

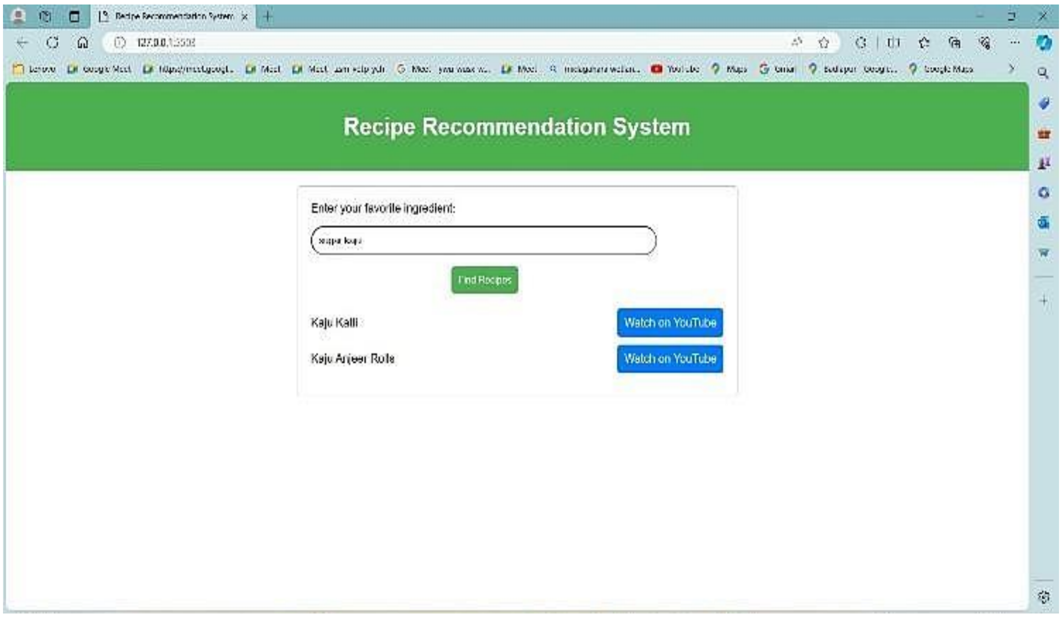

Figure 8. Recipe recommendation using the ingredients.

Figure 7 shows the feedback form which we have created for the user. After getting the ingredients and recipe for the image, which user provided for the classification, the user has to give the feedback for the provided recipe [21]. We are collecting the feedback for the purpose of knowing if users are satisfied with the recipe which model provides and also satisfied with the YouTube video which model provides for the recipe. This is the purpose of collecting the feedback form as shown in Figure 8.

DISCUSSION

We have also created one page which provides the recipe and the instruction video from YouTube for the understanding of the recipe. We created this page because if any user has some ingredients and wants to make any food then user can submit the ingredients and then it analyzes the ingredients and provides the recipe name and YouTube video for that recipe. For example, in Figure 8, the user gives the ingredient as a *kaju* (cashew nut) and sugar, then it provides the recipe name which will be made using these two ingredients as a *kaju Katli*. It also provides the YouTube link for that recipe.

CONCLUSION

The deep learning–based food recipe generation model has proven to be a promising and creative tool. The model successfully generated diverse and coherent recipes, showcasing adaptability to various cuisines and dietary preferences. Evaluation metrics such as perplexity, BLEU (BiLingual Evaluation Understudy) score, and recipe similarity affirmed the model's effectiveness in predicting word sequences and aligning with human-written recipes. While the model demonstrated creativity and user-friendly language, there are opportunities for improvement in capturing finer cooking details. Ongoing efforts in fine-tuning, user feedback integration, and potential expansion of the training dataset will contribute to enhancing the model's overall performance. Overall, this study marks a positive stride in leveraging deep learning for culinary creativity, laying the groundwork for future developments in artificial intelligence–driven recipe generation. The user-friendly outputs and promising results encourage further exploration and refinement of this technology, bringing us closer to a more sophisticated and user-centric approach to generating delectable recipes.

REFERENCES

1. Chen JJ, Ngo CW, Chua TS. Cross-modal recipe retrieval with rich food attributes. In: Proceedings of the 25th ACM International Conference on Multimedia, Mountain View, CA, USA, October 23–27, 2017. pp. 1771–1779.

2. Bettadapura V, Thomaz E, Parnami A, Abowd GD, Essa I. Leveraging context to support automated food recognition in restaurants. In: 2015 IEEE Winter Conference on Applications of Computer Vision, Waikoloa, HI, USA, January 5–9, 2015. pp. 580–587.
3. Salvador A, Drozdzal M, Giró-i-Nieto X, Romero A. Inverse cooking: recipe generation from food images. In: 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition, Long Beach, CA, USA, June 15–20, 2019. pp. 10453–10462.
4. Teng CY, Lin YR, Adamic LA. Recipe recommendation using ingredient networks. In: Proceedings of the 4th Annual ACM Web Science Conference, June 22–24, Evanston, IL, USA, 2012. pp. 298–307.
5. Zuo S, Xiao Y, Chang X, Wang X. Vision transformers for dense prediction: a survey. Knowledge-Based Syst. 2022; 253: 109552.
6. Salvador A, Drozdzal M, Giró-i-Nieto X, Romero A. Inverse cooking: recipe generation from food images. In: 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Long Beach, CA, USA, June 15–20, 2019. pp. 10453–10462.
7. Han F, Guerrero R, Pavlovic V. CookGAN: meal image synthesis from ingredients. In: 2020 IEEE/CVF Winter Conference on Applications of Computer Vision, Snowmass, CO, USA, March 1–5, 2020. pp. 1450–1458.
8. Fujita J, Sato M, Nobuhara H. Model for cooking recipe generation using reinforcement learning. In: 2021 IEEE 37th International Conference on Data Engineering Workshops (ICDEW), Chania, Greece, April 19–22, 2021. pp. 1–4.
9. Salvador A, Hynes N, Aytar Y, Marin J, Ofli F, Weber I, Torralba A. Learning cross-modal embeddings for cooking recipes and food images. In: 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, HI, USA, July 21–26, 2017. pp. 3020–3028.
10. Chhikara P, Chaurasia D, Jiang Y, Masur O, Ilievski F. Fire: food image to recipe generation. In: 2024 IEEE/CVF Winter Conference on Applications of Computer Vision (WACV), Waikoloa, HI, USA, January 3–8, 2024. pp. 8184–8194.
11. Wang L, Li Q, Li N, Dong G, Yang Y. Substructure similarity measurement in Chinese recipes. In: Proceedings of the 17th International Conference on World Wide Web, Beijing, China, April 21–25, 2008. pp. 979–988.
12. Pan S, Dai L, Hou X, Li H, Sheng B. ChefGAN: Food image generation from recipes. In: Proceedings of the 28th ACM International Conference on Multimedia, Seattle, WA, USA, October 12–16, 2020. pp. 4244–4252.
13. Salvador A, Gundogdu E, Bazzani L, Donoser M. Revamping cross-modal recipe retrieval with hierarchical transformers and self-supervised learning. In: 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition, Nashville, TN, USA, June 20–25, 2021. pp. 15475–15484.
14. Chen J, Ngo CW. Deep-based ingredient recognition for cooking recipe retrieval. In: Proceedings of the 24th ACM International Conference on Multimedia, Amsterdam, The Netherlands, October 15–19, 2016. pp. 32–41.
15. Chen J, Pang L, Ngo CW. Cross-modal recipe retrieval: how to cook this dish? In: Amsaleg L, Guðmundsson G, Gurrin C, Jónsson B, Satoh S, editors. MultiMedia Modeling: 23rd International Conference, MMM 2017, Reykjavik, Iceland, January 4–6, 2017, Proceedings, Part I 23. Cham, Switzerland: Springer International Publishing; 2017. pp. 588–600.
16. Wang X, Kumar D, Thome N, Cord M, Precioso F. Recipe recognition with large multimodal food dataset. In: 2015 IEEE International Conference on Multimedia & Expo Workshops (ICMEW), Turin, Italy, June 29–July 3, 2015. pp. 1–6.
17. Zhu B, Ngo CW, Chen J, Hao Y. R^2GAN: cross-modal recipe retrieval with generative adversarial network. In: 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Long Beach, CA, USA, June 15–20, 2019. pp. 11477–11486.
18. Wang S, Gao H, Zhu Y, Zhang W, Chen Y. A food dish image generation framework based on progressive growing GANs. In: Wang X, Gao H, Iqbal M, Min G, editors. Collaborative Computing: Networking, Applications and Worksharing: 15th EAI International Conference, CollaborateCom 2019, London, UK, August 19–22, 2019, Proceedings 15. Cham, Switzerland: Springer International Publishing. pp. 323–333.

19. Martinel N, Foresti GL, Micheloni C. Wide-slice residual networks for food recognition. In: 2018 IEEE Winter Conference on Applications of Computer Vision (WACV), Lake Tahoe, NV, USA, March 12–15, 2018. pp. 567–576.
20. Salvador A, Hynes N, Aytar Y, Marin J, Ofli F, Weber I, Torralba A. Learning cross-modal embeddings for cooking recipes and food images. In: 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, HI, USA, July 21–26, 2017. pp. 3020–3028.
21. Zhang XJ, Lu YF, Zhang SH. Multi-task learning for food identification and analysis with deep convolutional neural networks. J Computer Sci Technol. 2016; 31 (3): 489–500.

