Lecture Notes in Networks and Systems 648

Ajith Abraham ·
Thomas Hanne · Niketa Gandhi ·
Pooja Manghirmalani Mishra ·
Anu Bajaj · Patrick Siarry *Editors*

Proceedings of the 14th International Conference on Soft Computing and Pattern Recognition (SoCPaR 2022)

Springer

Swasth: An Inverse Cooking Recipe Generation from Food Images

G. N. R. Prasad[1(✉)], Y. Sri Lalitha[2], Y. Gayatri[2], and B. Indira[1]

[1] Department of MCA, Chaitanya Bharathi Insitute of Technology, Gandipet, Hyderabad, India
{gnrp,bindira_mca}@cbit.ac.in

[2] Gokaraju Rangaraju Institute of Engineering and Technology, Hyderabad, India

Abstract. The three important for every one of us is food, cloth and house. We give the first priority to food. At the same in this fast generation making a food by spending so many hours in the kitchen is a tough task. People are using various food delivery apps like Zomoto, Swiggy, Uber eats and many to get their food. Many restaurants have started takeaways courts to parcel the food items. Behind of every meal there is a procedure involved in making of it. How to the know the process behind of it. The basic of aim of this paper is to provide a solution to the query. We are introducing an inverse cooking system named it as "Swasth". This system recreates cooking recipes for the given food image. This system provides the ingredients used and then also gives the cooking instructions. Our system uses a unique architecture to forecast ingredients as sets, modelling their relationships without enforcing any order, and then creates cooking directions while concurrently paying attention to the image and its predicted components. We thoroughly test the system on the massive Recipe 1 million dataset and demonstrate that we are able to obtain high quality recipes by utilising both image and ingredients. We also demonstrate that the system is able to produce more compelling recipes than retrieval-based approaches in terms of human judgement.

Keywords: Inverse cooking · CNN · Deep learning · Machine learning

1 Introduction

In this planet, every one including species needs food to survive. Food gives us a energy. The food fines our identity and culture. Especially, the youth puts their favourite food items in the social media to say it we ate this food. Recently, the Instagram post #food reads to at least 300 million posts;

We always like to see the pictures of the food. These pictures attracts us to eat. There is also a process involved for every dish and by looking the picture some of us gets a question what is recipe for this food item. To get a solution of this question, here the software provides a solution.

The three important for every one of us is food, cloth and house. We give the first priority to food. At the same in this fast generation making a food by spending so many

Swasth is a Hindi language word means tasty.

A. Abraham et al. (Eds.): SoCPaR 2022, LNNS 648, pp. 871–878, 2023.
https://doi.org/10.1007/978-3-031-27524-1_85

hours in the kitchen is a tough task. People are using various food delivery apps like Zomoto, Swiggy, Uber eats and many to get their food. Many restaurants have started take away courts to parcel the food items. Behind of every meal there is a procedure involved in making of it.

Over time, eating traditions and cooking culture have also changed. Nowadays, we frequently eat meals that have been prepared by other parties, such as takeaway, caterers, and restaurants, but traditionally, the majority of meals were produced at home. Because it is difficult to acquire detailed information on prepared meals, it is tough to know exactly what we ingest. We argue that pre-packaged meals must be used to infer the components and cooking instructions for inverted cooking systems. Over the past few years, significant progress has been achieved in visual recognition tasks, such as object identification, semantic segmentation, and natural picture classification. When compared to natural picture understanding, however, food recognition presents more difficulties since food and its components have a high interclass variability and exhibit significant deformations due to cooking.

Ingredients come in a variety of colours, shapes, and textures when they are prepared, which usually obscures their appearance. Additionally, extremely complex reasoning and background information are required for visual ingredient detection. Because of this, food identification is a problem for modern computer vision systems, requiring them to go beyond the apparent and consider prior knowledge to provide high-quality structured food preparation descriptions. Food classification has been a major focus of previous attempts to better understand food. However, in addition to being able to recognise the type of meal and its components, a system for full visual food identification should also be able to understand how the food was prepared.

In a cooked dish, ingredients frequently get obscured and arrive in a range of hues, shapes, and textures. Additionally, visual ingredient recognition needs very sophisticated reasoning and prior knowledge. Food identification therefore presents a challenge to the state-of-the-art computer vision systems, forcing them to go beyond the obvious and take into account past knowledge to allow high-quality structured food preparation descriptions.

Traditionally, most meals were made at home, but nowadays, we regularly eat meals made by other sources, such as takeout, caterers, and restaurants. It is difficult to know exactly what we consume because there is little access to specific information about prepared meals. We contend that inverse cooking systems, which can deduce ingredients and cooking directions from a prepared meal, are necessary. Outstanding advancements have been made in visual recognition tasks over the past few years, including object identification, semantic segmentation, and natural picture categorization. When compared to natural picture understanding, however, food recognition presents more difficulties since food and its components have a high interclass variability and exhibit significant deformations due to cooking.

2 Releated Work

Prior attempts to better comprehend food have mostly concentrated on classifying ingredients and foods. However, a system for thorough visual food recognition should be able to comprehend the cooking procedure as well as the sort of meal or its contents. The picture-to-recipe issue has traditionally been conceptualised as a retrieval job, where a recipe is retrieved from a given dataset based on the image similarity score in an embedding space. The amount and diversity of the datasets, as well as the calibre of the learnt embedding, have a significant impact on these systems' performance. It should come as no surprise that these systems fail if the static data does not contain a recipe that matches the picture query. The limitations of the current approach are that a system for thorough visual food recognition should be able to grasp the preparation procedure as well as the kind of meal and its contents.

The proposed system works like a training will be done using Convolutional neural network (CNN) with recipe details and images and this model can be used to predict recipe by uploading related images and we used 1 million recipe dataset and from this dataset we used 1000 recipes as training entire dataset with images will take lots of memory and hours of time train CNN model.

The advantages of adopting the suggested system are that it explores various attention methods to simultaneously reason about both modalities, and it presents an inverse cooking system that creates cooking instructions conditioned on a picture and its materials. We thoroughly examine ingredients as a list and a set, and we suggest a novel architecture for ingredient prediction that takes use of interdependencies between constituents without enforcing order. By using a user research, we illustrate the superiority of our suggested system over methods for retrieving recipes from images and show that ingredient prediction is in fact a challenging problem.

3 System Architecture

The above figure represents the recipe generation model. We extract image features e_t with the image encoder, parametrized by θ_L Ingredients are predicted by θ_L, and encoded into ingredient embeddings e_L with θ_E. The cooking instruction decoder, parametrized by θ_R generates a recipe title and a sequence of cooking steps by attending to image embedding's e_I , ingredient embedding's e_L (Fig. 1).

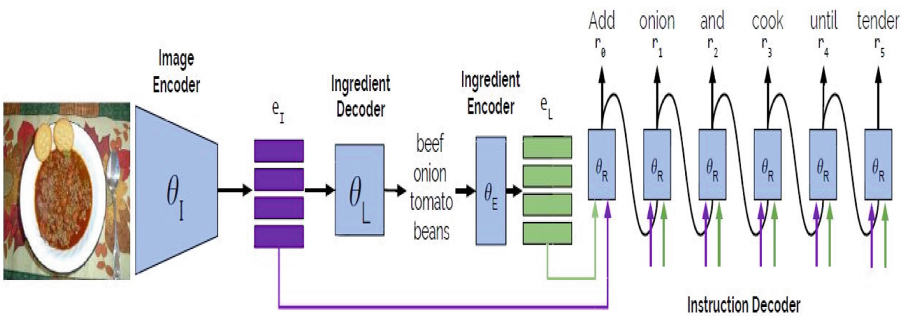

Fig. 1. System architecture [10]

4 Results

A convolutional neural network with 50 layers is called ResNet-50. The ImageNet database contains a pre trained version of the network that has been trained on more than a million photos. Word2vec embedding, bi-directional LSTM, and instruction retrieval tasks were employed for the ingredients extraction task. Our LSTM has two stages. We created a sample dataset and used several feature extraction techniques. It includes the files ingrs.pkl, instr.pkl, and demo pictures. Se-resnet-101 has a greater accuracy than resnet-50 and se-resnet50, according to the results (Fig. 2).

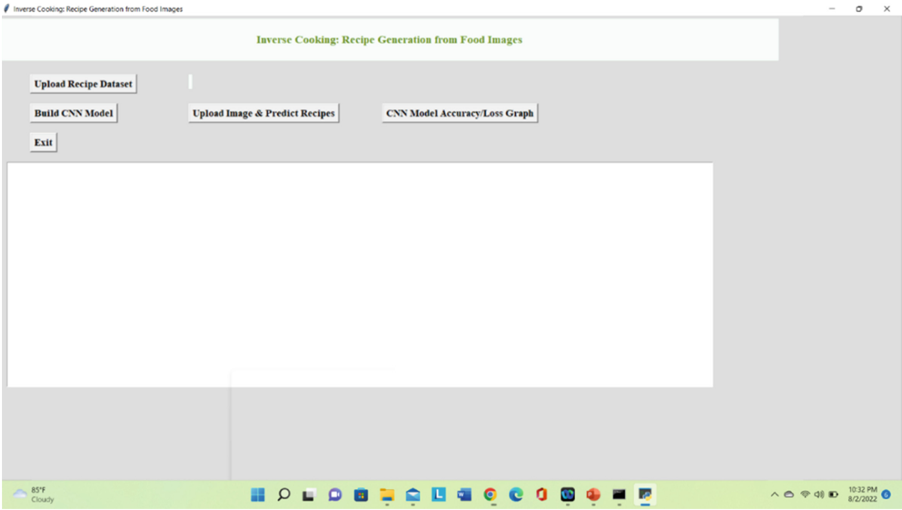

Fig. 2. User interface screen

Description: The above user interface allows the user to upload dataset using ‘Upload Recipe Dataset’ button (Fig. 3).

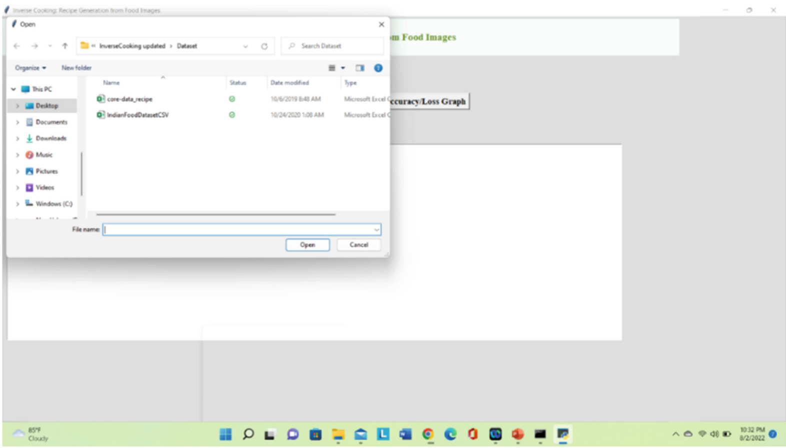

Fig. 3. Upload dataset

Description: In above screen selecting and uploading recipe dataset and then click on ‘Open’ button to load dataset and to get below screen (Fig. 4).

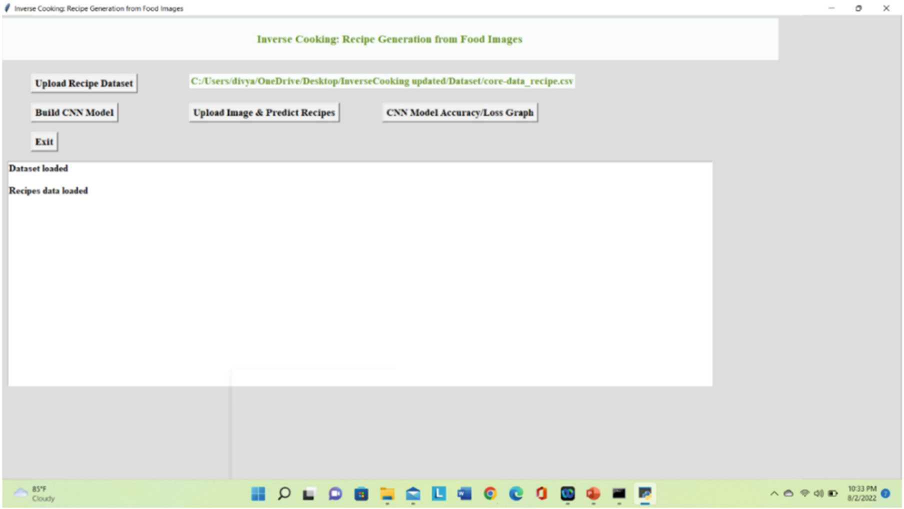

Fig. 4. Build machine learning model

Description: In above screen dataset loaded and now click on ‘Build CNN Model’ button to build CNN on above dataset (Fig. 5).

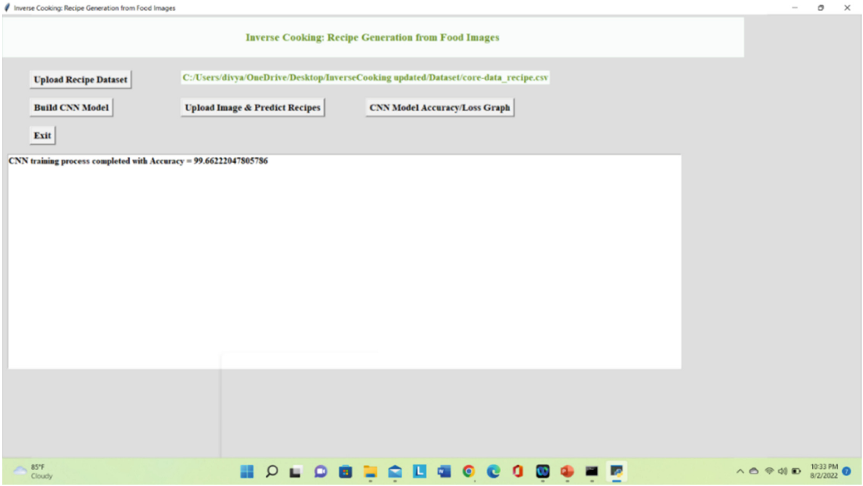

Fig. 5. Model built

Description: In above screen CNN model generated and we got prediction accuracy as 99.6%. Now click on 'Upload Image & Predict Recipes' button to upload test images (Fig. 6).

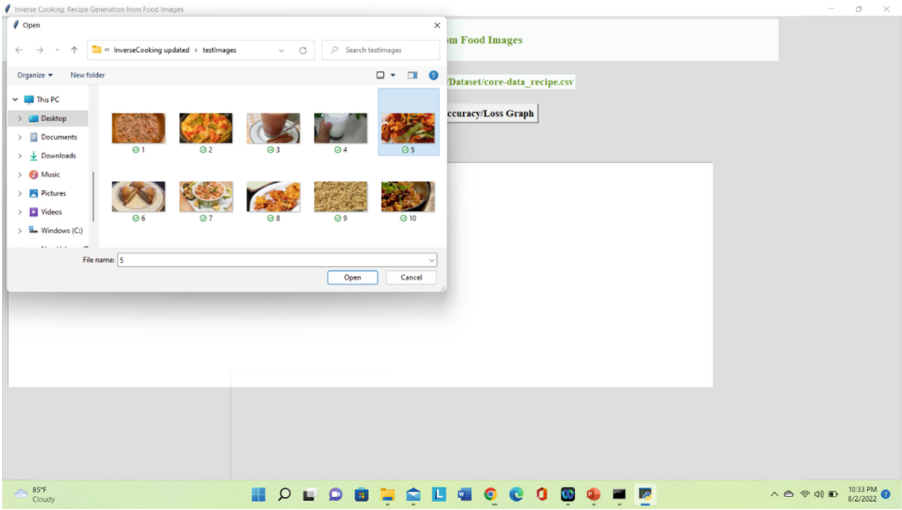

Fig. 6. Select test image

Description: In above screen select any image and then click on 'Open' button to get below result (Fig. 7).

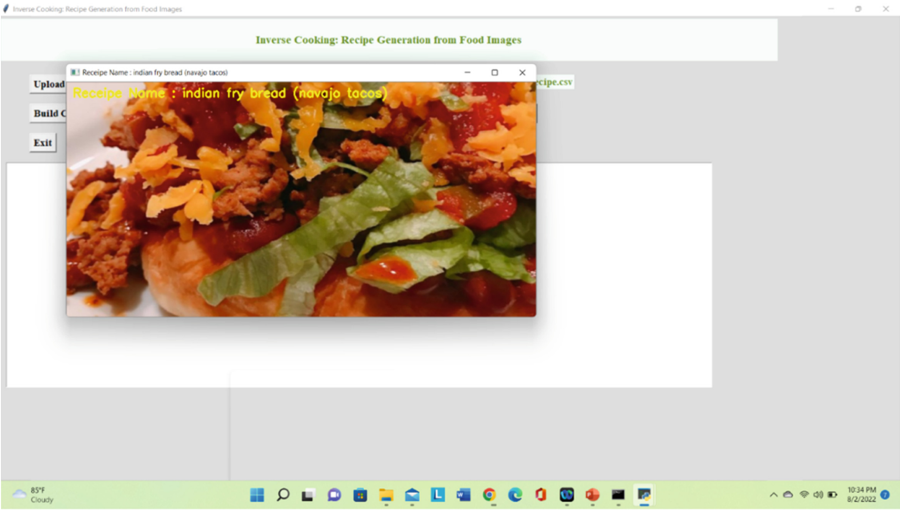

Fig. 7. Output

Description: In above screen uploaded image recipe identified as ‘indian fry bread(novojo tocos)’ and now close above image to get below details (Fig. 8).

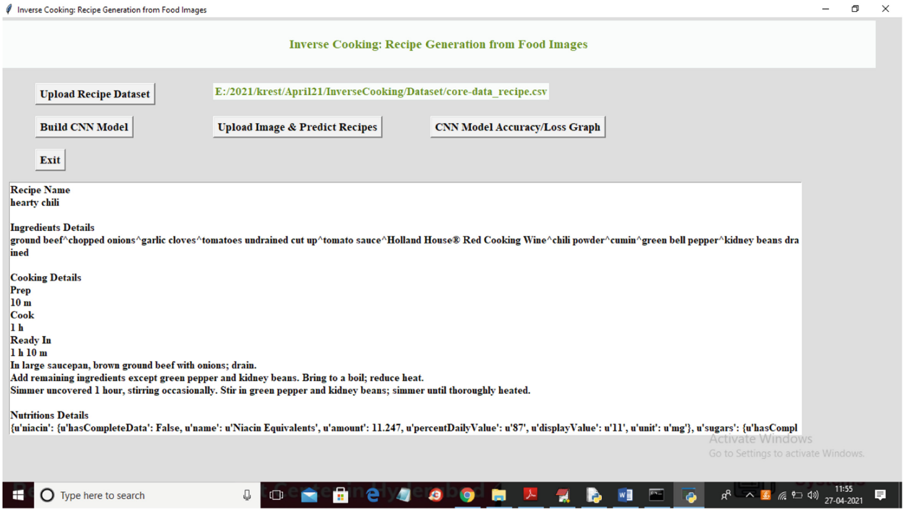

Fig. 8. Nutrition details

Description: In above screen we can see recipe name, ingredients details, cooking and nutrition details and similarly you can upload any image and get recipe (Fig. 9).

Now click on ‘CNN Model Accuracy/Loss Graph’ button to get CNN graph

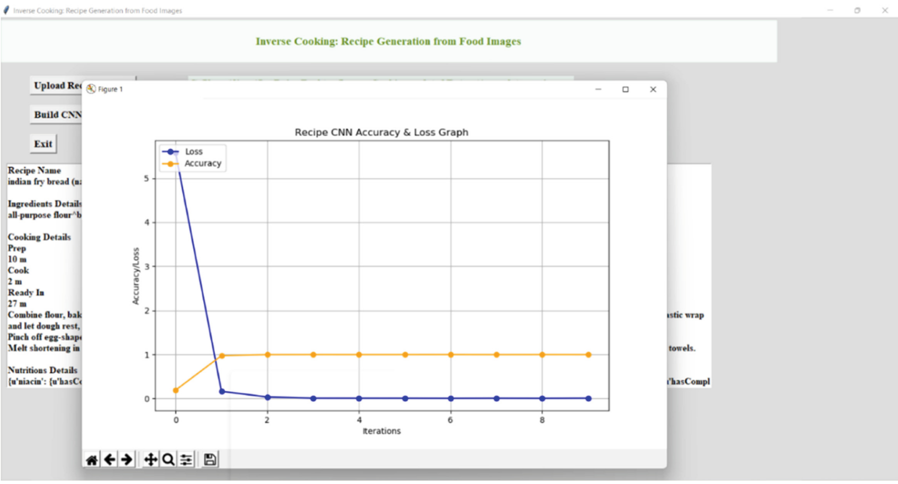

Fig. 9. Accuracy graph

Description: In above graph x-axis represents epochs and y-axis represents accuracy/loss value and blue line represents loss and orange line represents accuracy and in above graph with each increasing epoch accuracy got increase to 1 (100%) and loss decrease to 0. Any CNN model whose accuracy is high and loss is less will be consider as efficient model.

5 Conclusion

The suggested approach has great potential for usage in automatic recipe suggestions as well as information retrieval systems. With the use of an image of food, this method creates a recipe complete with a heading, list of components, and order of cooking instructions. We demonstrated the need of modelling dependencies by first predicting sets of elements from meal photos. The relevance of simultaneously inferring from both modes of reasoning was then highlighted when we looked into instruction generation conditioned on visuals and inferred constituents. Last but not least, user research findings support the task's complexity and show how better our system is to cutting-edge methods for retrieving recipes from images.

Declaration of Conflicting Interest. The author declares no conflicts of interest in the development of this research work.

Funding. This research receives no grant from any funding agency in the public, private, or not-for-profit sectors.

References

1. Bossard, L., Guillaumin, M., Van Gool, L.: Food-101 – mining discriminative components with random forests. In: Fleet, D., Pajdla, T., Schiele, B., Tuytelaars, T. (eds.) ECCV 2014. LNCS, vol. 8694, pp. 446–461. Springer, Cham (2014). https://doi.org/10.1007/978-3-319-10599-4_29
2. Carvalho, M., Cad'ene, R., Picard, D., Soulier, L., Thome, L., Cord, M.: Cross-modal retrieval in the cooking context: learning semantic text-image embeddings. In: SIGIR (2018)
3. Chen, J.-J., Ngo, C.-W.: Deep-based ingredient recognition for cooking recipe retrieval. In: ACM Multimedia. ACM (2016)
4. Chen, J.-J., Ngo, C.-W., Chua, T.-S.: Cross-modal recipe retrieval with rich food attributes. In: ACM Multimedia. ACM (2017)
5. Chen, M.-Y., et al.: Automatic chinese food identification and quantity estimation. In: SIGGRAPH Asia 2012 Technical Briefs (2012)
6. Chen, X., Zhou, H., Diao, L.: Chinesefoodnet: A large-scale image dataset for chinese food recognition. CoRR, abs/1705.02743 (2017)
7. Dai, B., Lin, D., Urtasun, R., Fidler, S.: Towards diverse and natural image descriptions via a conditional gan. ICCV (2017)
8. Dembczy´nski, K., Cheng, W., H¨ullermeier, E.: Bayes optimal multilabel classification via probabilistic classifier chains. In: ICML (2010)
9. Fan, A., Lewis, M., Dauphin, Y.: Hierarchical neural story generation. In: ACL (2018)
10. Fischler, C.: Food, self and identity. Information (International Social Science Council) (1988)