IJNRD.ORG ISSN : 2456-4184
INTERNATIONAL JOURNAL OF NOVEL RESEARCH AND DEVELOPMENT (IJNRD) | IJNRD.ORG
An International Open Access, Peer-reviewed, Refereed Journal

Identify Ingredients From Food Images And Generate Recipe

Guide: [1] Dr. Amrapali Chavan, Authors: [2] Ronit Patil, [3] Ninad Shirsat, [4] Kunal Bumb, [5] Atharva Satpute

[1] Faculty, Computer Engineering Department, AISSMS Institute of Information Technology, Pune, India
[2] Student, Computer Engineering, AISSMS Institute of Information Technology, Pune, India
[3] Student, Computer r Engineering, AISSMS Institute of Information Technology, Pune, India
[4] Student, Computer Engineering, AISSMS Institute of Information Technology, Pune, India
[5] Student, Computer Engineering, AISSMS Institute of Information Technology, Pune, India

Abstract. Identifying ingredients from food images and generating recipe problem involves the task of creating a recipe with a desired dish image as input. In recent years, there has been increasing interest in developing machine learning methods to solve this problem. This article presents a review of the current state of the art of generating recipes, focusing on the various techniques that have been proposed to solve this problem, including neural networks, probabilistic modelling and so on. and rules-based systems. The paper also discusses challenges that still need to be addressed in the field, such as the lack of large-scale datasets and the difficulty of modelling complex cooking techniques. Finally, the paper highlights several potential future directions for generating recipe research, such as incorporating additional sources of information, such as users' nutritional preferences and needs, into the model formula form.

***Keywords* -** Machine Learning, Semantic Integration, Data Augmentation, Probabilistic Modelling, Algorithm encoding.

A. INTRODUCTION

Reverse cooking is a fascinating subject of research aimed at creating a recipe for a certain dish. Although many people have cooking experience, reverse cooking requires the ability to understand the complex relationships between ingredients, cooking techniques, and flavours. Reverse cooking has many real-world applications, such as creating recipes for meal planning services, personalized recommendations for food delivery, and culinary education.

In recent years, there has been an increasing interest in using machine learning techniques to solve the reverse cooking problem. These techniques range from traditional rule-based systems to the most advanced neural network architectures. Although considerable progress has been made, there are still many challenges to be addressed in this area, such as the lack of large-scale data sets, the difficulty of modelling complex cooking techniques, and the need to Create personalized recipes.

In this context, reverse cooking offers an exciting opportunity to combine culinary knowledge with machine learning techniques. By doing so, we are able to develop intelligent systems capable of generating high-quality recipes that meet individual preferences and dietary requirements. The field is changing rapidly and this is an exciting time to get involved in reverse cooking system research and development.

- Lack of standardized recipe datasets.

One of the main challenges of reverse cooking is the lack of a standardized recipe dataset. The recipe datasets used to train machine learning models can vary widely in quality, consistency, and format. This can cause problems with the accuracy and generalizability of the resulting models. Additionally, some recipe datasets may favor certain cuisines or cooking styles, limiting the model's applicability to other dishes. To address this, researchers are working to develop more comprehensive and standardized recipe datasets, as well as techniques for cleaning and validating existing datasets.

- Limited availability of training dataset.

Another major challenge of reverse cooking is the limited training data available. This is especially true for niche or regional cuisines, which may have fewer recipes. Therefore, it can be difficult to form precise patterns for these types of dishes. Additionally, even for the most popular dishes, the number of recipes available can still be limited, which can lead to overfitting and other model performance issues. Researchers are exploring various techniques to tackle this problem, such as data augmentation and knowledge transfer from related dishes.

- Variation in cooking methods.

Cooking methods can vary widely, even within the same dish or recipe, which can be a challenge for reverse cooking models. Different cooking methods can lead to significant differences in the taste, texture, and other characteristics of the final dish, which can make it difficult to generalize to patterns for recipes. new cooking or cooking style. In addition, some cooking methods may not be explicitly represented in the recipe dataset, which further limits the models' ability to learn and generalize. Researchers are exploring various approaches to solving this problem, such as integrating more detailed information about cooking methods into models or using multimodal learning techniques to capture different aspects of the cooking process.

- Lack of domain knowledge.

Domain knowledge is an important factor in inverse cooking, as it can help improve the accuracy and generalizability of models. For example, chefs and other culinary professionals can provide valuable insights into the cooking process, the interactions between ingredients, and other factors that can affect the final dish. However, incorporating domain knowledge into machine learning models can be challenging, as it often involves encoding complex qualitative information in a way that algorithms can use. To address this, researchers are exploring techniques such as knowledge graphs, semantic integration, and other methods that can help capture and leverage domain knowledge in cooking models in reverse.

II. METHODOLOGY

Producing a formula from a picture may be a challenging assignment, which needs a synchronous understanding of the fixings composing the dish as well as the changes they went through, for example, slicing, mixing or blending with other fixings. Instead of getting the formula from a picture straightforwardly, we contend that a formula era pipeline would advantage from a halfway step foreseeing the fixings list.

The arrangement of instructions would at that point be created conditioned on both the picture and its comparing list of fixings, where the interaction between picture and fixings seem provide additional experiences on how the last mentioned were prepared to deliver the coming about dish. Figure 4 outlines our approach. Our formula era framework takes a nourishment picture as an input and yields a sequence of cooking informational, which are created by means of an instruction decoder that takes as input two embeddings. The primary one speaks to visual highlights extracted from an picture, whereas the moment one encodes the fixings extricated from the picture.

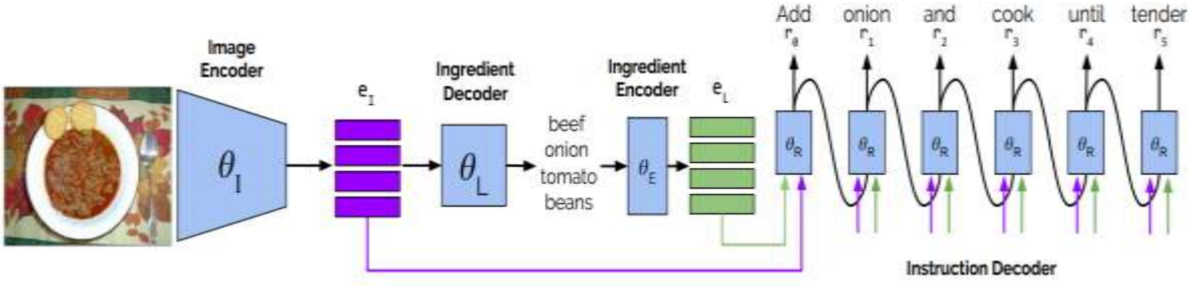


Fig. 4 recipe generator core model

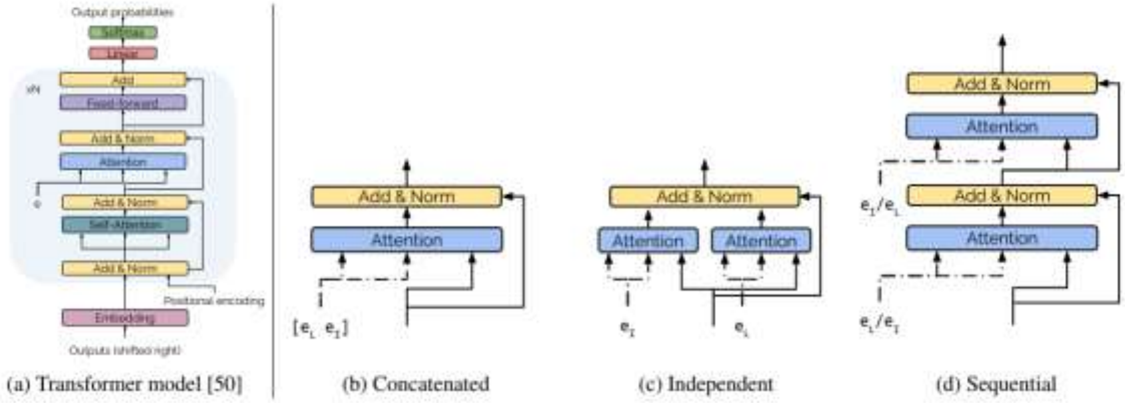
If the input is an image are then we the goal is to create a sequence of commands R = (r_1, ..., r_T) (where rt means word in order) using instruction transformer. Note that the title is predicted as the first instruction. This transformer is connected to two inputs: image representation e_I and an ingredient containing e_L. We took pictures representation by the ResNet-50 coder and get ingredient embedding using the e_L decoder architecture to predict ingredients followed by a single embedding a layer that maps each component to a vector of fixed size. The command decoder consists of a transformer blocks, each containing two attentional layers followed by a linear layer. The first layer of attention applies self-attention to previously created results, while the other ensures that the model is in good condition to specify the self-check printout.

Transformer model consists of several transformation blocks followed by a linear layer and SoftMax nonlinearity that gives the distribution of recipe words for each time step t. Figure 3a illustrates a transformer model conventionally conditioned to a single mode. However, our recipe generator relies on two sources: image properties $e_I \in R^{P\times s}$ and components $e_L \in R^{K\times de}$ (P and K represent the number of image and items, respectively, and de is the delivery dimension). That is why we want to draw our attention to the reasons for both categories that simultaneously control the command generation process. To do this, we explore three different fusions strategies:

- **Combined attention**. This strategy first combines both image and constituent e_L embeddings over the first dimension $e_{concat} \in R^{(K+P)\times de}$. Then attention is paid to combined closures.

- **Independent attention.** This strategy includes bilevel attention to procedural bimodal conditioning. In this case, one layer is on top of the image embedding e_I, while the other participates in gradient embeddings e_L. Get the attention of both the layers are combined using a merge operation.

- **Sequential attention.** This strategy involves two conditioning methods in sequence. In our design, we consider two orders: (1) image first where attention is first calculated based on the embedding of images e_I and then component attachments e_L; and (2) ingredients first when the order is placed and us first participate in e_L input of ingredients and then e_I input of images.



(a) Transformer model [50]
(b) Concatenated
(c) Independent
(d) Sequential

Fig. 5 instruction decoder attention strategy. In our experiments, we used the attention module as transformer (a) with three different attention modules (b-d) for generating cooking instructions using multiple conditions.

III. RELATED WORK

Understanding food. Large scale deployment food data such as Food-101 and Recipe1M, together with the recent iFood Challenge[2] made possible significant advances in visual food recognition providing training benchmarks and comparing machine learning methods. As a result, there is now a vast computer vision literature dealing with various on food-related tasks, especially in image classification. Later jobs deal with more complex tasks such as evaluation the number of calories given in the food image by evaluating the food amounts predicting the current ingredient list and find a recipe for a given image. In addition, provides a detailed cross-regional analysis of food recipes taking into account images, features (e.g. style and flow) and recipe ingredients. Tasks related to food has also been addressed in the natural language processing literature where recipe generation has been studied context to create procedure text from both streams' tables or ingredient checklists.

Classification of several signs. Considerable efforts have been made in the literature dedicated to the use of deep neural networks for multi-character classification through model design and the study of loss functions are well suited this task. Early companies use a single label classification models combined with binary logistic loss assuming that independence between tags and removal of potentially relevant information. One way to capture label dependencies is to rely on label power series. Power sets consider all possible sticker combinations, making them hard to figure out for large problems. Another expensive option is to learn the joint probability of the labels. Probabilistic classification chains are used to solve this problem. and their recurrent neural network-based counterparts suggest collapsing the joint distribution into conditional expressions at the expense of introducing internal order. Note that most of these models require the prediction of every possible identifier. In addition, a joint contribution and label constructions were carried out maintain correlations and predict sets of labels. Alternatively, researchers have tried to predict cardinality a set of tags; Assuming, however, that the labels are independent. About the multi-character classification targets, binary logistic loss, cross entropy of the target distribution, mean squared error of the target distribution and rank-based losses were studied and in comparison.

Recent results from large-scale data sets illustrate target distribution loss potential. Conditional text generation. Generating conditional text using autoregressive models has been widely studied both literature-based and text-based in image-based conditions. In neural machine translation that aims to predict Various architectural models have been studied to translate a given source text into another language, including recurrent neural networks, convolutional models and attentional approaches. More recently, sequence-to-sequence models have been applied to more open model's generational tasks such as poetry and storytelling. Following the trends in neural machine translation, autoregressive models have shown promising performance inscriptions, where the goal is briefly describing the content of the image, opening doors to less limited problems such as manufacturing descriptive paragraphs or visual storytelling.

IV. EXPERIMENTS

Train and evaluate a model with Food 101; This dataset contains 101 categories of food, with 1000 images per category, for a total of 101,000 images. The dataset also includes the corresponding recipes for each image. The dataset contains 8,974 training sessions. 32,036 verified and 34,045 tested recipes include title, list of ingredients, list of cooking instructions, and (Optional) Image. The dataset was obtained by scraping a cooking website (www.foodspotting.com), so the resulting recipes are very unstructured and often contain redundant or very narrowly defined dishes. Ingredients (for example olive oil, virgin olive oil, Spanish olive oil. Oil is another ingredient). All images can be found in the "images" folder and are organized per class. All image ids are unique and correspond to the foodspotting.com review ids. The test/train splitting used in the experiment of our paper, can be found in the "meta" directory. The dataset structure is as follows:

```
pec/
        images/
                    <class_name>/
                    <image_id>.jpg
        meta/
                    classes.txt
                    labels.txt
                    test.json
                    test.txt
                    train.json
                    train.txt
```

Overall, we reduce our vocabulary of over 15,000 ingredients to about 2,000 unique ingredients. For cooking guides, tokenize the raw text, remove words that occur less than 10 times in the dataset, Replace them with unknown word tokens. besides us Also add special tokens at the beginning and end of the recipe as the end of the lesson. This process leads to the recipe A vocabulary of about 22,000 unique words.

Resize the shortest side of the image to 256 pixels, take a random 224×224 crop for training and select Centre 224 x 224 pixels for evaluation. The command decoder uses a 16-block, 8-block transformer. Attention with multiple heads, each with dimension 64. As a component decoder, we use a transformer with four blocks. and two multi-headed attentions, each with dimensionality Use the last convolutional layer of the ResNet-50 model to obtain the image embeddings. The embedding dimensions for both the image and the component are 512. we hold the maximum Use 20 ingredients per recipe and follow directions 150 words maximum. The model is trained by Adam the optimizer waits until an early stopping criterion is met (using a patience of 50 and watching for validation loss). all models are Implemented with PyTorch. Additional implementation details are included in the supplementary material.

In this section, we compare the proposed constituent prediction methods with previously implemented models the goal is to assess whether ingredients should be processed such as lists or sets. We consider models from several brands Break the literature down into entry levels and match them to our goals. On the other hand, we have food-based models feedforward convolutional networks trained to predict the components. Let's try several losses to train these models, namely binary cross entropy, soft distribution connection, and target distribution cross entropy. Note that binary cross entropy is the only one that does not exist considering the dependencies between a set of elements. On the other hand, we have sequential models that predict list, sort and exploit dependencies elements.

Finally, let's review the recently proposed models each pair determined the cardinality of the prediction by prediction determines which elements to include. Table 1 (right) shows the results of the validation set for top-level basic settings as well as recommended settings is approaching, we evaluated the model using points of union (IoU) and F1 scores calculated for the cumulative number of T P, F N and F P in the entire data set separation (according to Pascal's VOC convention). As in the picture in the table, the feedforward model trained with binary cross entropy (FF_{BCE}) has the lowest performance both measures that can be explained by the assumed independence of the components. These results already exist significantly improves the method that learns to predict the set cardinal (FF_{DC}). Likewise, performance increases when training a model with structured losses such as soft IoU (FF_{IOU}). Our future model is trained with a target distribution (FF_{TD}) and a threshold value (th = 0.5) the sum of selected constituent probabilities exceeds all input-output levels, including recently proposed alternatives to ensemble forecasting such as (FF_{DC}). Note that target distribution models depend on example array elements and implicitly captures the cardinality information.

We follow the model series found in the latest literature as lists, we train a transformer network for component prediction is given by minimizing the negative log likelihood loss (TF list).

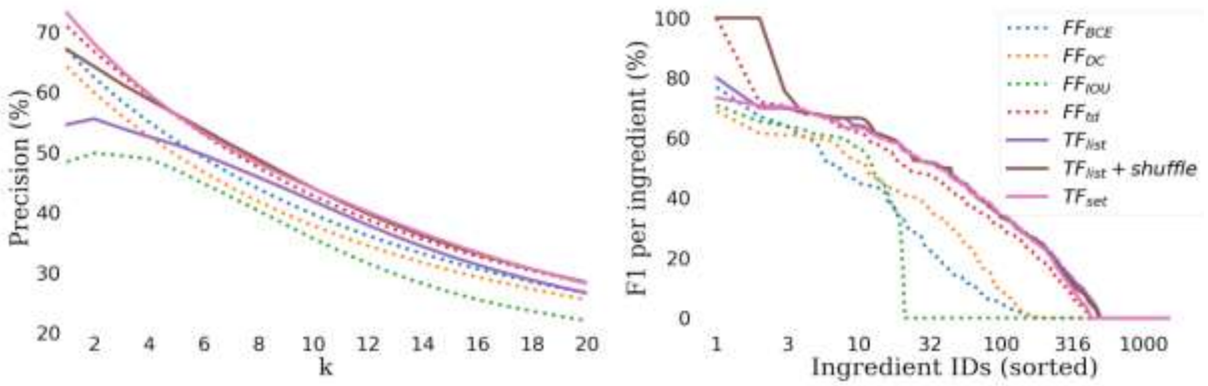


Fig. 6 P vs K and F1 – score per ingredient

V. CONCLUSION

We recommend modelling recipes with cooking programs. We designed a program outline and annotated a set of recipes programs and we presented an approach to learn to predict programs from food images and recipes. Experimental results show that the projectionCommon space between images and recipes in the program can improve recovery results. Finally, we showed how can we create images of food by controlling a program. We hope that the programs will open up new directions such as like allowing dealers to run recipes or allowing us to extract common sense knowledge about foods from graphs. Limits and social impact. In this work, we do not beyond predicting the cooking process through the program.

However, nutritional value prediction, calorie estimate and their impact on our health is an important topic. In addition, incorrect prediction schedules may not be possibly make or result in inedible food. In addition, component prediction models should be applied consciously, especially among users. food allergy cases. Future work involves an analysis about potential biases that our training programs or data may have (for example, for unhealthy dishes, for Western world with underrepresented cuisines) and its impact on the food industry.

VI. REFERENCES

[1] Michael Carvalho, Rémi Cadène, David Picard, Laure Soulier, Nicolas Thome and Matthieu Cord. 2018. Cross-Modal Retrieval in the Cooking Context: Learning Semantic Text-Image Embeddings. In Proceedings of the 41st International ACM SIGIR Conference on Research and Development in Information Retrieval, Ann Arbor, MI, USA (SIGIR'18). ACM, New York, NY, USA

[2] Food Image to Cooking Instructions Conversion Through Compressed Embeddings Using Deep Learning Madhu Kumari Tajinder Singh Computer Science and Engineering Department Computer Science Engineering Department National Institute of Technology Chandigarh University Hamirpur, H.P., INDIA Punjab, INDIA, madhu.jaglan@gmail.com nith2kl4@gmail.com

[3] The creation of recipes using food images Salvador, Amaia Mikel Drozdzal Tom´as Xavier Giro-i-Nieto Ingrid Romero University of Catalonia for Technical Studies Facebook AI Research may be reached at adrianars, mdrozdzal, and maia.salvador@upc.edu

[4] RECIPEGM: A Hierarchical Recipe Generation Model Anja Reusch, Alexander Weber, Maik Thiele, and Wolfgang Lehner Database Systems Group Technische Universitat Dresden ¨ Dresden, Germany anja.reusch, alexander arno.weber, maik.thiele, wolfgang.lehner@tu-dresden.de

[5] TYPICALITY ANALYSIS OF INGREDIENT COMBINATION IN A COOKING RECIPE FOR HELPING WITH INGREDIENT ARRANGEMENT Graduate School of Information Science, Nagoya University, Japan. Satoshi Yokoi1, Keisuke Doman2, Takatsugu Hirayama1, Ichiro Ide1, Daisuke Deguchi3, and Hiroshi Murase1. School of Engineering, Chukyo University, Japan, yokois@murase.m.is.nagoya-u.ac.jp, Hirayama Ide, murase@is.nagoya-u.ac.jp, ddeguchi@nagoya-u.jp

[6] BRS bench: A Recipe for Cooking Benchmarking Program One Frederic André's National Institute of Informatics' Research Division for Digital Content and Media Sciences Japan, Tokyo, Chiyoda-ku, 2-1-2 Hitotsubashi

[7] Based on the primary ingredients and primary seasoning, user-generated recipe sites may be clustered to find similar recipes. Akiyo Email address: nadamoto@konan-u.ac.jp Nadamoto Konan University Okamoto 8-9-1 Higashinada-ku Kobe, Japan Email: m1424007@center.konan-u.ac.jp Shunsuke Hanai Konan University Okamoto 8-9-1 Higashinada-ku Kobe, Japan Namba Hidetsugu Email: nanba@hiroshima-cu.ac.jp Hiroshima City University Ozukahigashi 3-4-1 Hiroshima 731-3194 Japan.

[8] Structure-Aware Generation Network for Image-Based Recipe Generation Steven C. H. Hoi, Chunyan Miao, Guosheng Lin, and Hao Wang School of Computer Science and Engineering, Nanyang Technological University, Singapore Singapore Management University Joint NTU-UBC Research Center of Excellence in Active Living for the Elderly, NTU, gslin,ascymiao@ntu.edu.sg chhoi@smu.edu.sg

[9] RECIPES GENERATED USING DEEP LEARNING FROM FOOD IMAGES Srinivasamoorthy. Dr. Preeti Savant PG student at JAIN University in Karnataka, India, department of computer applications Assistant professor at JAIN University in Karnataka, India's Department of Computer Application

[10] Marc Bravin, Lucerne School of Information Technology, Lucerne University of Applied Sciences and Arts, 6343 Rotkreuz, Switzerland, Deep Inverse Cooking VM 02 Master of Science in Engineering

[11] GourmetNet: Multi-Scale Waterfall Features with Spatial and Channel Attention for Food Segmentation Andreas Savakis, Bruno Artacho, and Udit Sharma

[12] A Cooking Recipes Dataset for Semi-Structured Text Generation is called RecipeNLG. Micha Bien', Micha Gilski', Martyna Maciejewska', Wojciech Taisner', Dawid Wisniewski', and Agnieszka Lawrynowicz' all have positions at the faculty of computing and telecommunications at the Poznan University of Technology in Poland.

[13] Understanding and Using a Cooking Robot to Follow Recipes Tyler Thompson, Nicholas Roy, Mario Bollini, Stefanie Tellex, and Daniela Rus

[14] Recipe Generation for Inverse Cooking from Food Photographs and Cuisine Grouping Ruby K. oza1 I'm Aanal S. Raval1. I'm Aakanksha V. Jain1. Ahmedabad's Silver Oak College of Engineering and Technology's Department of Computer Engineering, M.E. India's Gujarat

[15] "Food Image Analysis and Recipe Generation: A Review" by L. Gao et al. (2020).

[16] "Recipe Generation from Food Images: A Deep Learning Approach" by M. Han et al. (2020).

[17] "Recipe Generation from Food Images using Attention-based Neural Networks" by D. Chaudhary et al. (2020).

[18] "Food Recognition and Recipe Generation from Food Images: A Review" by P. Sharma et al. (2019).

[19] "Recipe Generation from Food Images using Deep Neural Networks" by H. Kim et al. (2019).

[20] "Recipe Generation from Food Images using Deep Learning and NLP" by Y. Liu et al. (2019).