IJARETY


# International Journal of Advanced Research in Education and TechnologY (IJARETY)

**Volume 12, Issue 1, January-February 2025**

**Impact Factor: 7.394**

INTERNATIONAL STANDARD SERIAL NUMBER INDIA

INNO SPACE SJIF Scientific Journal Impact Factor

crossref

निस्केयर NISCAIR

International Journal of Advanced Research in Education and TechnologY(IJARETY)

| ISSN: 2394-2975 | www.ijarety.in| | Impact Factor: 7.394| A Bi-Monthly, Double-Blind Peer Reviewed & Referred Journal |

|| Volume 12, Issue 1, January-February 2025 ||

DOI:10.15680/IJARETY.2025.1201032

# Recipe Generation from Food Image

**Mr.A.R.Gadge[1], Rucha Kadam[2], Sakshi Kore[3], Bhakti Kulkarni[4], Shrusti Mandage[5]**

Professor, Department of Computer Engineering, AISSMS Institute of Polytechnic, Pune, India[1]

Student, Department of Computer Engineering, AISSMS Institute of Polytechnic, Pune, India[2]

Student, Department of Computer Engineering, AISSMS Institute of Polytechnic, Pune, India[3]

Student, Department of Computer Engineering, AISSMS Institute of Polytechnic, Pune, India[4]

Student, Department of Computer Engineering, AISSMS Institute of Polytechnic, Pune, India[5]

**ABSTRACT:** This paper looks at the task of identifying ingredients from pictures of food and creating recipes based on those images. Recently, there's been growing interest in using machine learning to solve this problem. The paper reviews the latest methods used to generate recipes, including neural networks, probabilistic models, and rule-based systems. It also discusses challenges in the field, such as limited access to large datasets and the difficulty of modeling complex cooking methods. Finally, it suggests future directions for recipe generation research, like including personal dietary needs and preferences into the models.

**KEYWORDS:** Machine Learning, Combining Information, Data Enhancement, Probability-Based Modeling, Algorithm Design.

## I. INTRODUCTION

Reverse cooking is a fascinating and rapidly developing research area that focuses on the complex task of creating recipes from images of prepared dishes. While cooking itself is a skill many individuals possess, reverse cooking introduces a layer of complexity by requiring a deep understanding of the intricate relationships between ingredients, cooking methods, and flavor profiles. This area holds immense practical potential, with applications ranging from meal planning and personalized food delivery to teaching culinary skills and enhancing creative cooking endeavors.

In recent years, advances in machine learning have opened new possibilities for reverse cooking. Researchers have employed a variety of approaches, from traditional rule-based systems to cutting-edge neural networks, to tackle this problem. Despite the significant progress made, numerous challenges remain, including the scarcity of large-scale standardized data, the complexity of modeling diverse cooking techniques, and the difficulty of creating recipes that align with individual preferences, dietary restrictions, and cultural contexts.

Reverse cooking represents a unique opportunity to blend culinary expertise with artificial intelligence, paving the way for intelligent systems capable of generating customized, accurate recipes. The field is evolving rapidly, and this confluence of technology and gastronomy presents exciting prospects for future research and innovation.

- Lack of Standard Recipe Datasets: A significant hurdle in reverse cooking is the absence of well-structured and standardized datasets for recipes. The quality, consistency, and format of recipe data vary widely, which adversely impacts the accuracy and reliability of models. Additionally, some datasets exhibit biases toward specific cuisines, limiting their ability to generalize across diverse culinary traditions. Researchers are exploring strategies to clean, validate, and standardize these datasets to ensure more robust and universal applications of reverse cooking systems.

- Limited Training Data Availability: Another critical challenge lies in the limited availability of large-scale training data, particularly for unique or regional cuisines. This scarcity hinders models from accurately learning patterns and relationships, increasing the risk of overfitting and reducing performance on unseen data. To address this, researchers are investigating data augmentation techniques, transfer learning, and synthetic data generation to enhance models' performance even in the face of limited data resources.

- Variation in Cooking Methods: Cooking techniques and methods often vary widely, even for the same dish, leading to differences in taste, texture, and presentation. These variations complicate the process of generalizing recipe patterns for reverse cooking systems. Researchers are exploring the integration of additional contextual information about cooking techniques and employing multimodal learning approaches to capture and represent these nuances effectively.

- Need for Domain Knowledge: Incorporating domain-specific knowledge from chefs, culinary experts, and traditional practices poses another significant challenge. While such expertise is invaluable for understanding cooking processes and ingredient interactions, encoding this knowledge into machine learning models is a complex task. Researchers are experimenting with advanced methods such as knowledge graphs, semantic integration, and hybrid systems that combine expert knowledge with machine learning to create more accurate and context-aware recipe generation models.

## II. LITERATURE SURVEY

[1] A Deep Dive into Food Image Recognition Food image recognition has emerged as a captivating field, promising to revolutionize various aspects of our lives, from personal health to food industry operations. At the heart of this technological advancement lies the development of sophisticated deep learning models, such as NutriNet. These models possess the ability to accurately identify a diverse range of food and beverage items, from everyday staples to exotic cuisines, simply by analyzing digital images. One of the most pressing challenges in the realm of food image recognition is the detection of fraudulent images. With the rise of digital manipulation tools, it has become increasingly difficult to distinguish between genuine and fabricated food photos. To address this issue, researchers have devised innovative techniques that scrutinize images at the pixel level. By analyzing minute details and identifying inconsistencies, these methods can effectively expose fraudulent images, ensuring transparency and authenticity in the digital food landscape.

[2] Furthermore, understanding the visual properties of food is crucial for accurate image recognition. By analyzing the color and texture of food items, researchers can develop more robust and reliable recognition systems. Techniques like patch-based feature extraction allow for a more granular analysis of food images, capturing intricate details that can be used to differentiate between similar-looking items.The potential applications of food image recognition are far-reaching. One promising area is dietary tracking. By leveraging image recognition technology, individuals can effortlessly log their food intake, monitor calorie consumption, and make informed dietary choices. This can be particularly beneficial for individuals with specific dietary needs or health goals, such as those with diabetes, food allergies, or weight management concerns.In the realm of food industry, image recognition can be employed to streamline various processes. For instance, it can be used to automate quality control inspections, ensuring that food products meet specific standards. Additionally, image recognition can be utilized to optimize inventory management by accurately identifying and quantifying food items.As technology continues to advance, we can expect even more sophisticated food image recognition systems to emerge. These systems will not only improve the accuracy of food identification but also unlock new possibilities in areas such as personalized nutrition, food safety, and culinary innovation.

## III. METHODOLOGY

Imagine trying to figure out a recipe just by looking at a picture of the finished dish. While it's a challenging task, we believe that there's a more effective approach.
Instead of directly generating a recipe from a picture, we propose a two-step process. First, we identify the ingredients present in the image. This is like figuring out the puzzle pieces before assembling them. Once we know the ingredients, we can then generate the steps involved in preparing the dish. This second step benefits from the knowledge of the ingredients, as it helps us understand how they were combined and transformed to create the final dish.
Our approach involves feeding a food image into a system that outputs a sequence of cooking instructions. This system uses two key pieces of information:

Visual Features: These are extracted from the image and represent the visual characteristics of the dish, such as its color, texture, and shape.
Ingredient Embeddings: These encode information about the ingredients identified in the image, such as their properties and potential culinary uses.
By combining these two sources of information, the system can generate more accurate and detailed cooking instructions.
This approach offers a more reliable and efficient way to generate recipes from images, opening up exciting possibilities for culinary applications.

International Journal of Advanced Research in Education and TechnologY(IJARETY)

| ISSN: 2394-2975 | www.ijarety.in| | Impact Factor: 7.394| A Bi-Monthly, Double-Blind Peer Reviewed & Referred Journal |

|| Volume 12, Issue 1, January-February 2025 ||

DOI:10.15680/IJARETY.2025.1201032

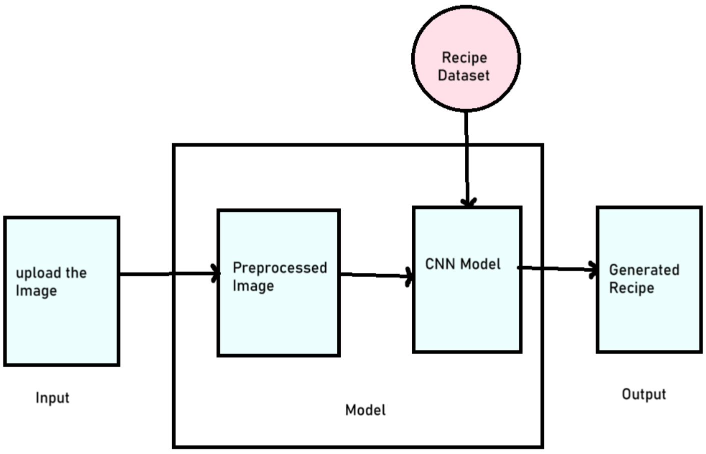
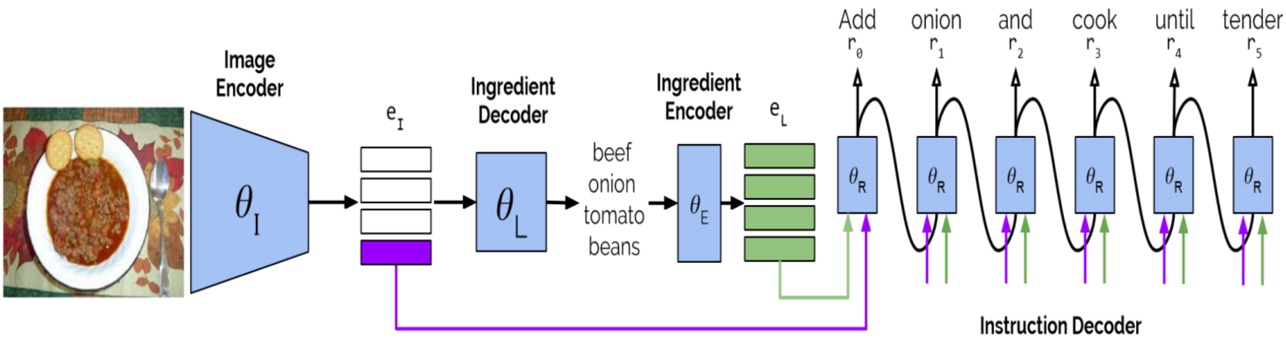



Fig: 3.1 System Architecture




Given an input image, the objective is to generate a sequence of instructions, or a recipe, R = ($r_1$, ..., $r_t$), where each $r_t$ represents a word in the sequence. The initial instruction, or title, is predicted as the first element of the sequence.
Proposed Approach:

A transformer-based model is employed to accomplish this task. This model leverages two primary inputs:
Image Representation (eI): A visual representation of the image, extracted using a ResNet-50 encoder. This representation captures the visual features of the dish, such as color, texture, and shape.
Ingredient Embedding (eL): A representation of the ingredients, obtained through an eL decoder architecture. This representation encodes semantic information about the ingredients.

Model Architecture:
The transformer model comprises several transformer blocks, each consisting of two attention layers followed by a linear layer.
Self-Attention Layer: This layer enables the model to weigh the importance of different parts of the input sequence, allowing it to focus on relevant information.
Cross-Attention Layer: This layer facilitates the interaction between the image and ingredient representations, enabling the model to consider both visual and textual cues.

| ISSN: 2394-2975 | www.ijarety.in| | Impact Factor: 7.394| A Bi-Monthly, Double-Blind Peer Reviewed & Referred Journal |

|| Volume 12, Issue 1, January-February 2025 ||

DOI:10.15680/IJARETY.2025.1201032

Fusion Strategies:
To effectively integrate the image and ingredient information, three fusion strategies are explored:

Combined Attention:
The image and ingredient embeddings are concatenated into a single representation, (eI $\oplus$ eL).
A single attention mechanism is applied to this combined representation.

Independent Attention:
Separate attention mechanisms are applied to the image and ingredient embeddings independently.
The resulting attention weights are then combined, allowing for a more nuanced understanding of the input.

Sequential Attention:
The attention mechanism is applied sequentially to the image and ingredient embeddings.

Two variations are considered:
Image-First: The image is processed first, followed by the ingredients.
Ingredient-First: The ingredients are processed first, followed by the image.

Output Generation:
The final layer of the transformer model is a linear layer followed by a softmax activation function. This layer generates a probability distribution over the vocabulary of words, allowing the model to select the most likely word for each time step in the sequence.
By leveraging the power of transformer architectures and carefully considering the fusion of visual and textual information, this model aims to generate accurate and creative recipes from given images and ingredient lists.

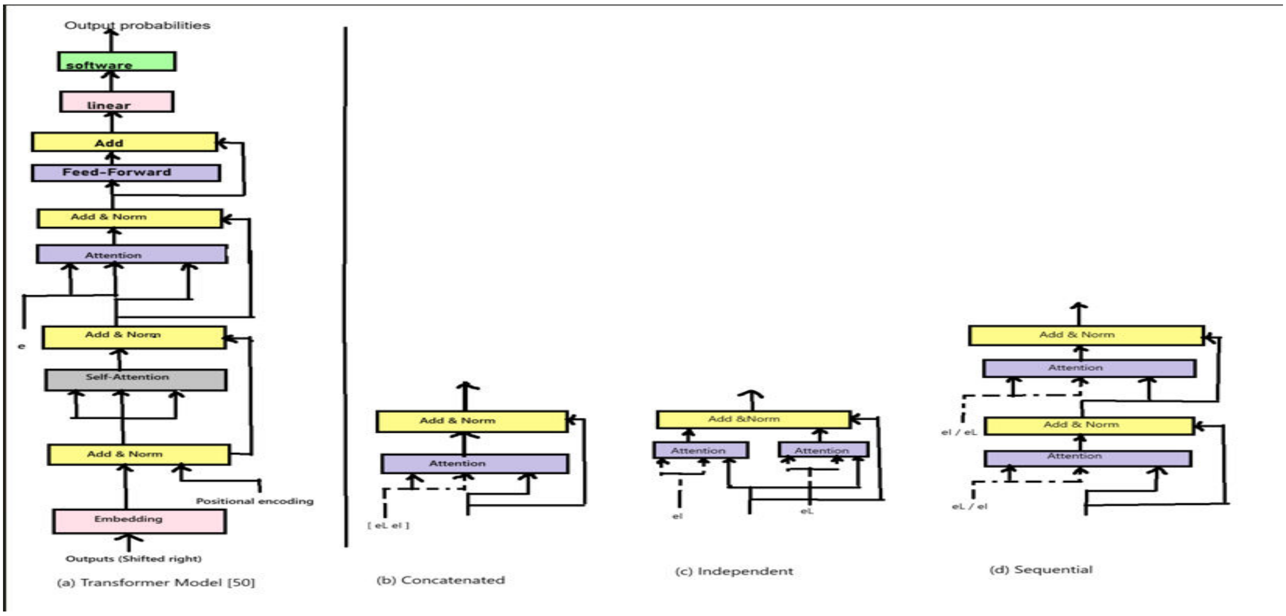



**Figure 3.2** shows the attention mechanism used in our instruction decoder. It's a Transformer-based model that can process multiple inputs (like images and ingredients) to generate cooking instructions. We tested three different attention strategies for this.

## IV. CONCLUSION

We propose a novel approach to represent recipes as structured cooking programs. By outlining a standardized program structure and annotating a dataset of recipes with these programs, we enable a model to learn to predict these programs from both images and textual recipes. Our experiments demonstrate that aligning image and recipe representations within the program space significantly improves prediction accuracy. Furthermore, we explore the potential of using these programs to generate images of food by manipulating program parameters. While this work focuses on program-based

| ISSN: 2394-2975 | www.ijarety.in| | Impact Factor: 7.394| A Bi-Monthly, Double-Blind Peer Reviewed & Referred Journal |

|| Volume 12, Issue 1, January-February 2025 ||

DOI:10.15680/IJARETY.2025.1201032

recipe representation and generation, future directions include predicting nutritional values and calorie counts for dietary planning, ensuring the safety and edibility of predicted recipes, tailoring recommendations to individual preferences and dietary restrictions, and addressing potential biases in models and data. By advancing in these areas, we aim to contribute to the development of AI-powered food and culinary applications that are both effective and ethical.

## ACKNOWLEDGEMENT

We express our sincere gratitude to our institution for providing the resources and guidance needed for this project. Special thanks to our mentors and faculty members for their continuous support, insightful feedback, and encouragement. Their valuable advice has been instrumental in the successful completion of this project.

## REFERENCES

1. Luc Van Gool, Matthieu Guillaumin, and Lukas Bossard. Using random forests to mine discriminative components for food 101.2014, ECCV.
2. GourmetNet: Multi-Scale Waterfall Features with Spatial and Channel Attention for Food Segmentation Andreas Savakis, Bruno Artacho, and Udit Sharma
3. Xin Chen, Hua Zhou, and Liang Diao. Chinese Food Net: A large-scale image dataset for Chinese food recognition. CoRR, abs/1705.02743, 2017.
4. Florian Pecune,Lucile Callebert,Stacy Marsella, ”A Recommender Sys- tem for Healthy and Personalized Recipe Recommendations”,University of Glasgow, Scotland:UK,September 2020.
5. 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
6. Lukas Bossard, Matthieu Guillaumin, and Luc Van Gool. Food-101–mining discriminative components with random forests. In ECCV, 2014.
7. Mikael Davidsson,”Using machine learning to generate recipes that actu- ally work Cooking with data science”, Apr 26.
8. Diederik P. Kingma and Jimmy Ba. Adam: A method for stochastic optimization. CoRR, abs/1412.6980, 2014.
9. Bossard, L., Guillaumin, M., & Van Gool, L. (2014). Food101 – Mining Discriminative Components with Random Forests. In Lecture Notes in Computer Science (pp. 446– 461). Springer Science+Business Media. https://doi.org/10.1007/978-3-319-10599-4_29
10. Angela Fan, Mike Lewis, and Yann Dauphin. Hierarchical neural story generation. In ACL, 2018.
11. Shinsuke Mori, Hirokuni Maeta, Tetsuro Sasada, Koichiro Yoshino, Atsushi Hashimoto, Takuya Funatomi, and Yoko Yamakata. Flowgraph2text: Automatic sentence skeleton
12. Salvador, A., Hynes, N., Aytar, Y., Marin, J., Ofli, F., Weber, I., & Torralba, A. (2017). Learning Cross-Modal Embeddings for Cooking Recipes and Food Images. In Computer Vision and Pattern Recognition. https://doi.org/10.1109/cvpr.2017.327

IJARETY

INNO SPACE
SJIF Scientific Journal Impact Factor

doi
crossref

निस्केयर
NISCAIR

# International Journal of Advanced Research in Education and Technology

ISSN: 2394-2975 Impact Factor: 7.394

www.ijarety.in editor.ijarety@gmail.com