IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 5, Issue 6, February 2025
Impact Factor: 7.67

Recipe Generation from Food Images Using Deep Learning

Siddhi Gharat[1] and Tanmay Tandel[2]
Assistant Professor Department of IT[1]
Student, P.G. Department of IT[2]
Veer Wajekar ASC College, Phunde, Uran

Abstract: *Food imagery has become a dominant mode of content sharing on digital platforms, inspiring interest in automatic food recognition and recipe generation. This research explores the possibility of converting a food image into a usable cooking recipe—including a list of ingredients and detailed preparation instructions—using deep learning. A comprehensive system is proposed that integrates convolutional neural networks (CNNs) for image feature extraction, multi-label classification for ingredient detection, and transformer-based models for instruction generation. The system leverages inverse cooking models and attention mechanisms to produce coherent and accurate recipes. Evaluation of the model shows an overall recipe generation accuracy of 86%, indicating the potential of deep learning in culinary applications and smart kitchen systems.*

Keywords*:* Deep learning, food recognition, recipe generation, image-to-text, inverse cooking, neural networks

I. INTRODUCTION

Cooking is a fundamental human activity that encompasses cultural, economic, and health dimensions. The rise of social media and food-centric platforms like Instagram, Pinterest, and food blogs has led to a proliferation of food images. These images present an opportunity to convert visual food content into actionable formats such as recipes, thereby enhancing user engagement, assisting in meal planning, and supporting personalized dietary recommendations. Traditionally, recipes are created manually or generated through structured datasets. However, recent advances in computer vision and natural language processing (NLP) have opened up the possibility of generating recipes automatically from images. This paper presents a reverse cooking system that analyzes a food image to generate a complete recipe. The proposed architecture is based on CNNs for image feature extraction, followed by multi-label classification for ingredient prediction, and sequence-to-sequence models using transformers for cooking instruction generation.

II. LITERATURE REVIEW

A substantial body of work has explored the application of deep learning to food image analysis and recipe generation:

Srinivasamoorthy et al. (2022) proposed a system for recipe generation from food images, highlighting challenges such as visual variation and intra-class similarity. Their work emphasizes multi-modal learning for better accuracy.**Salvador et al. (CVPR 2019)** introduced the concept of inverse cooking, combining CNNs and RNNs to infer ingredients and generate textual instructions from food images. Their model integrates image embeddings with natural language generation.

Gao et al. (2020) provided a comprehensive review on food image analysis, discussing various methods, datasets, and challenges in automatic recipe generation. **Han et al. (2020)** proposed a deep learning-based method for generating recipes, discussing architecture design, dataset preparation, training, and performance evaluation.

Chaudhary et al. (2020) explored attention-based neural networks for recipe generation, showing improvements in capturing contextual relationships among ingredients and instructions. **Sankar et al. (2021)** examined broader applications of AI in cooking, including culinary creativity, meal planning, and personalized assistance. **Ujwala et al.**

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429 Volume 5, Issue 6, February 2025 Impact Factor: 7.67

(2023) developed a CNN-based inverse cooking system that avoids predefined processing order by using attention mechanisms to predict ingredient dependencies and generate stepwise instructions.

III. RESEARCH GAP

Despite promising developments, current systems face several limitations: Traditional models like NutriNet rely heavily on hand-crafted features (color, texture), limiting their generalization. Personalized classifiers assume clean and structured data, which is often unavailable in real-world scenarios.

Inverse Cooking models are sensitive to the quality of image-text embeddings, impacting performance. These challenges underline the need for more robust, end-to-end systems capable of handling diverse and noisy inputs with greater accuracy.

IV. METHODOLOGY

The proposed reverse cooking system consists of a three-stage pipeline:

4.1 Image Feature Extraction

A pre-trained Convolutional Neural Network (CNN), such as ResNet-50 or DenseNet201, is used to encode the food image into a high-dimensional feature vector. The CNN is fine-tuned on food-specific datasets to capture intricate visual patterns.

4.2 Ingredient Inference

The encoded image features are input to a multi-label classification model that predicts a binary vector representing the presence or absence of ingredients. The loss function used is Binary Cross Entropy (BCE):

$BCE = -1/N \sum [y_i * \log(p_i) + (1 - y_i) * \log(1 - p_i)]$

Here, y_i is the ground truth label for ingredient i, and p_i is the predicted probability.

4.3 Instruction Generation

A transformer-based encoder-decoder architecture is used to generate textual instructions. The encoder processes the image features and predicted ingredient embeddings, while the decoder generates text using attention mechanisms. Beam search is employed during decoding to generate fluent and contextually appropriate instructions.

V. SYSTEM ARCHITECTURE

5.1 User Input Module

- **Image Upload**: Users can upload a food image via a web interface.
- **Text Search**: Alternatively, users can enter a dish name to retrieve a predefined recipe.

5.2 Image Processing Module

- **Preprocessing**: Image is resized, normalized, and passed through a pre-trained CNN for feature extraction.
- **Encoding**: Features are encoded into a NumPy array for efficient similarity comparison.

5.3 Similarity Matching

- **Cosine Similarity** is used to match the input image features with a dataset of pre-encoded food images.
- **Recipe Retrieval**: Top matches are used to fetch corresponding recipes from the database.

5.4 Recipe Generation Module

- **Ingredient Prediction**: A multi-label classification model outputs likely ingredients.
- **Instruction Generation**: A transformer-based model generates the cooking steps.

5.5 Output Interface

- **Display**: Recipes, including ingredients, cooking steps, preparation time, and optional garnishing tips, are displayed to the user.

ISSN: 2581-9429 Volume 5, Issue 6, February 2025 Impact Factor: 7.67

VI. RESULTS

The model was trained and evaluated using a benchmark dataset of food images annotated with ingredients and instructions. Key results include:

- **Ingredient prediction accuracy**: 82%
- **Instruction BLEU Score**: 0.67
- **Overall recipe generation accuracy**: 86%

The system performed well on common and well-photographed dishes. However, performance dipped slightly for complex or visually ambiguous dishes, highlighting the need for further fine-tuning.

VII. DISCUSSION

The results indicate strong potential for real-world deployment:

- **Applications**: Cooking assistants, smart kitchen appliances, diet trackers.
- **Challenges**: Handling occluded or mixed dishes, detecting hidden ingredients, scaling to diverse cuisines.
- **Opportunities**: Integration with voice assistants, regional language support, and nutritional estimation.

VIII. CONCLUSION

This paper demonstrates the feasibility of recipe generation from food images using an integrated deep learning framework. By combining CNN-based image encoding, multi-label classification, and transformer-based instruction generation, the proposed system achieves high accuracy and user-friendliness. Future enhancements could include real-time feedback, personalization features, and augmented reality integration for guided cooking.

REFERENCES

[1]. Salvador, A., et al. (2019). Inverse Cooking: Recipe Generation From Food Images, CVPR.
[2]. Chen, J., & Ngo, C. (2016). Deep-based Ingredient Recognition for Cooking Recipe Retrieval, ACM Multimedia.
[3]. Bossard, L., et al. (2014). Food-101 – Mining Discriminative Components, Springer LNCS.
[4]. Chen, M., et al. (2012). Automatic Chinese Food Identification and Quantity Estimation, ACM SIGGRAPH.
[5]. Srinivasamoorthy, S., et al. (2022). Recipe Generation from Food Images using Deep Learning.
[6]. Gao, L., et al. (2020). Food Image Analysis and Recipe Generation: A Review.
[7]. Chaudhary, D., et al. (2020). Recipe Generation from Food Images using Attention-based Neural Networks.
[8]. Sankar, S., et al. (2021). Cooking with AI: A Survey on Recipe Generation using Deep Learning.
[9]. Ujwala, B., et al. (2023). Inverse Cooking Recipe Generation from Food Images