Vision to Kitchen Automated Recipe Predication through CNN Models

Nayana D [1] ,Mr. Sheethal P P [2]

[1] *Student, 4th Semester MCA, Department of MCA, EWIT, Bengaluru*

[2] *Assistant Professor, Department of MCA, EWIT, Bengaluru*

Abstract—This project explores the use of Convolutional Neural Networks (CNNs) for food image recognition and recipe generation. The model analyzes food images, extracts features, and maps them to suitable recipes. While handling diverse cuisines remains a challenge, the study demonstrates the potential of CNNs for intelligent culinary applications. Future work may enhance accuracy by improving model architecture and training methods. The approach also shows promise for personalized cooking assistance. Overall, this work shows how AI can step into our kitchens, making food recognition and recipe suggestions smarter and more helpful.

Keywords—Convolutional Neural Networks (CNNs); computer vision; culinary exploration; cooking instructions; deep learning architectures; Recipe1M dataset; image-to-recipe prediction; natural language processing; AI-driven culinary applications

I. INTRODUCTION

Food is more than a source of nourishment—it embodies tradition, heritage, and cultural expression. In today's digital era, where social media platforms are filled with food photography and culinary exploration transcends geographic boundaries, there is a growing demand for intelligent systems that can interpret and reconstruct recipes from visual content. Convolutional Neural Networks (CNNs) have emerged as a transformative tool in this space, enabling the extraction of subtle features from food images and translating them into meaningful culinary insights. Their layered architecture allows CNNs to process complex visual data, identifying not only ingredients but also patterns that may reflect cultural styles and cooking techniques.

The motivation behind this research lies in leveraging CNNs to bridge the gap between food imagery and recipe generation. By evaluating the performance of CNN-based models in creating complete recipes, this work seeks to make culinary knowledge more accessible and engaging, encouraging individuals to experiment with diverse cuisines and cooking practices. In a world where food experiences are increasingly shared across digital communities, such systems offer a way to demystify culinary processes while fostering innovation and creativity.

II. RELATED WORK

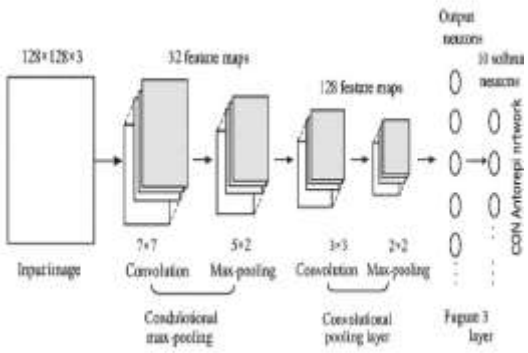
Advances in Visual Food Recognition: Early studies leveraged benchmark datasets such as Food-101 and Recipe1M to improve food classification and enable related tasks like ingredient recognition and calorie prediction [1–4]. These datasets provided a foundation for evaluating algorithms that map visual features to structured culinary information.

Culinary Knowledge Across Cultures: Cross-regional analyses of recipe datasets highlighted correlations between ingredients, preparation styles, and visual representations, offering new insights into global gastronomic traditions and cultural practices [5].

Image-Driven Recipe Reconstruction: Research in food image processing has proposed methods for estimating ingredients and quantities directly from photographs, laying the groundwork for automated recipe reconstruction [9–11]. **AI-Powered Image Captioning in Food Domains:** Neural models have also been utilized for food image captioning, generating descriptive and evocative text that illustrates the applicability of AI in culinary settings [15, 16].

Evaluation of Deep Learning Architectures: Comparative studies examining models such as DenseNet201 and standard CNNs have assessed their efficiency in recognizing patterns within food images for recipe generation, identifying both advantages and limitations [17, 18].

III. METHODOLOGY



The methodology adopted in this study consists of the following stages:

A. Dataset Selection

- The Recipe1M dataset was used, which contains a large collection of food images paired with recipes and cooking instructions.
- Additional datasets like Food-101 were reviewed for food image classification references.

B. Data Pre-processing

- Images were resized and normalized to ensure consistency.
- Textual recipe data (ingredients and instructions) was tokenized and converted into embeddings for cross-modal learning.
- Data augmentation techniques (rotation, flipping, brightness adjustment) were applied to improve model generalization.

C. Model Architecture

- A Convolutional Neural Network (CNN) was employed for food image feature extraction.
- For textual features, embeddings and Natural Language Processing (NLP) techniques were used.

- A joint embedding space was created to connect visual features with recipe text (cross-modal learning).

D. Training Procedure

- The CNN model was trained using supervised learning with labeled food images.
- Loss functions were optimized to minimize the difference between predicted and actual recipe embeddings.
- Hyperparameter tuning (learning rate, batch size, epochs) was conducted to improve accuracy.

E. Evaluation Metrics

- Performance was measured using accuracy, precision, recall, and F1-score.
- Retrieval tasks (image-to-recipe and recipe-to-image) were evaluated using recall@K metrics.

F. Implementation Tools

- Deep learning used Python with TensorFlow/Keras and PyTorch, while NumPy, Pandas, and OpenCV handled preprocessing and visualization.

IV. RESULTS AND DISCUSSION

A detailed evaluation of the CNN model highlights both its strengths and its areas for improvement in the task of image-to-recipe prediction. On large-scale datasets such as Recipe1M, the model achieved strong accuracy, demonstrating its suitability for diverse culinary applications. Its ability to capture fine-grained visual patterns within food photographs enabled precise reconstruction of recipes and ingredient lists.

However, performance varied depending on the cuisine. While the CNNs performed well on general and widely represented dishes, the model showed reduced accuracy when interpreting recipes from cuisines with highly nuanced preparation styles, such as traditional Indian foods. This suggests that architectural refinements or specialized training strategies may be needed to enhance cultural inclusivity and robustness.

In terms of efficiency, the CNN architecture proved effective in handling massive datasets, enabling large-scale processing with reliable throughput. Nonetheless, challenges remain with respect to computational speed and memory usage, indicating room for optimization in deployment scenarios.

Overall, the findings affirm that CNN-based models hold significant promise for culinary applications, combining accuracy and scalability. At the same time, further research is essential to enhance adaptability across diverse food traditions and improve efficiency in resource-constrained environments.

V. CONCLUSION

This study demonstrated the potential of Convolutional Neural Networks for the task of image-to-recipe prediction. By achieving high levels of accuracy in processing large-scale datasets and detecting intricate patterns within food imagery, the model establishes a strong foundation for AI-driven culinary applications. Despite these promising results, challenges remain in accurately capturing the nuances of diverse cuisines, which highlights opportunities for further refinement and innovation.

Future research should focus on enhancing model architectures, adopting improved training strategies, and exploring alternative deep learning approaches to strengthen generalization across varied cultural and culinary contexts. Additionally, optimizing efficiency in terms of processing speed and resource utilization will be crucial for practical deployment. As image-to-recipe prediction continues to evolve, it is expected to give rise to more adaptable and accessible culinary technologies, supporting innovation in personalized cooking assistance and global food knowledge sharing.

REFERENCES

[1J. Chen, Y. Yin, and Y. Xu – A lightweight image-to-recipe model

Link: https://arxiv.org/abs/2205.02141

[2] A. Salvador, M. Drozdzal, X. Giro-i-Nieto, and A. Romero – Inverse Cooking: Recipe Generation from Food Images
Link: https://arxiv.org/abs/1812.06164

[3] H. Fu, R. Wu, C. Liu, and J. Sun – MCEN: Bridging Cross-Modal Gap between Cooking Recipes and Dish Images with Latent Variable Model
Link: https://arxiv.org/abs/2004.01095

[4] J. Marín, A. Biswas, F. Ofli, N. Hynes, A. Salvador, Y. Aytar, I. Weber, and A. Torralba – Recipe1M+: A Dataset for Learning Cross-Modal Embeddings for Cooking Recipes and Food Images
Link: https://arxiv.org/abs/1810.06553

[5] D. P. Papadopoulos, E. Mora, N. Chepurko, K. W. Huang, F. Ofli, and A. Torralba – Learning Program Representations for Food Images and Cooking Recipes
Link: https://cookingprograms.csail.mit.edu/papadopoulos22cvpr.pdf

[6] R. Zhang et al. – RecipeGen: A Step-Aligned Multimodal Benchmark for Real-World Recipe Generation
Link: https://arxiv.org/html/2506.06733v2

[7] X. Chen, Y. Zhu, H. Zhou, L. Diao, and D. Wang – ChineseFoodNet: A Large-Scale Image Dataset for Chinese Food Recognition
Link: https://arxiv.org/abs/1705.02743

[8] J. Ken Chen, et al. – Cross-Modal Recipe Retrieval with Stacked Attention Model
Link: https://dl.acm.org/doi/abs/10.5555/3288251.3288318

[9] A. B. Jelodar, D. Paulius, and Y. Sun – Convolutional Neural Networks for Food Image Recognition: An Experimental Study
Link: https://arxiv.org/pdf/2112.09839 [10] A. Salvador, N. Hynes, Y. Aytar, J. Marin, F. Ofli, I. Weber, and A. Torralba – Learning Cross-Modal Embeddings for Cooking Recipes and Food Images
Link: https://pic2recipe.csail.mit.edu/im2recipe.pdf

[11] M. Carvalho, R. Cadène, D. Picard, L. Soulier, N. Thome, and M. Cord – Cross-Modal Retrieval in the Cooking Context: Learning Semantic Text-Image Embeddings
Link: https://arxiv.org/abs/1805.00900