IARJSET ISSN (O) 2393-8021, ISSN (P) 2394-1588

International Advanced Research Journal in Science, Engineering and Technology

Impact Factor 8.066 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 11, Issue 5, May 2024

DOI: 10.17148/IARJSET.2024.11589

Food Image Pattern Recognition and Recipe Generation Using Convolutional Neural Networks

Champakamala S [1], Bhoomika B U [2], Harshitha L Solanki [3]

Asst. Prof, Dep of ISE, Sri Siddhartha Institute of Technology, Tumkur, India [1]

UG Student, ISE Dept, Sri Siddhartha Institute of Technology, Tumkur, India [2]

UG Student, ISE Dept, Sri Siddhartha Institute of Technology, Tumkur, India[3]

Abstract*:* Computer science is experiencing rapid growth. Reconstructing cooking recipes from photos of food presents an intriguing challenge. The goal is to create complete recipes with ingredient lists, titles, and comprehensive instructions using convolutional layers in CNNs. About identifying complex patterns in food photos, this study clarifies the capabilities and limitations of CNNs by assessing critical performance metrics like recipe generation accuracy and efficiency. This work helps the culinary industry develop new technological solutions in response to the increasing need for a thorough understanding of meal preparation. Moreover, the implications of this research go beyond computer science, as it has the potential to drastically change how people interact with food. This research opens the door to a more diverse and interconnected culinary landscape by democratizing culinary knowledge and fostering a deeper understanding of global gastronomic traditions. In the end, this study's conclusions will guide future research into creating AI-powered culinary apps that are customized to each user's unique preferences and tastes, enhancing the appreciation of cooking around the world.

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

It becomes clear as we explore the world of food that a dish is more than just its ingredients; it also represents a tradition, history, and cultural legacy. In a time when social media is a thriving place for food photography and where culinary exploration is boundless, there is an increased need for in-depth knowledge of the craft of cooking. Convolutional Neural Networks (CNNs) are the cornerstones of innovation that make it possible to understand the subtleties concealed in food photos. Because of their hierarchical architecture, CNNs can transform complex visual data into insightful understandings that allow for a more nuanced understanding of the culinary landscape. In addition to identifying ingredients, this study aims to utilize CNNs' capacity to interpret cultural influences and cooking technique nuances.

This research aims to democratize culinary knowledge by analyzing CNNs' performance in creating recipes, giving people the confidence and creativity to take on culinary adventures. The culinary world is a patchwork of tastes, textures, and cultural influences, and every dish reflects the creativity of people throughout history.

The need for tools to solve the mysteries of food imagery has never been greater than in the digital age when food experiences are shared and celebrated worldwide. CNNs are the innovators in this field, providing a means of deciphering the intricacies hidden in food photos. This research aims to democratize culinary knowledge by assessing CNNs' performance in creating recipes, spurring a fresh wave of culinary innovation and exploration. CNNs' potential lies in helping us explore uncharted territories.

IARJSET ISSN (O) 2393-8021, ISSN (P) 2394-1588

International Advanced Research Journal in Science, Engineering and Technology

Impact Factor 8.066 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 11, Issue 5, May 2024

DOI: 10.17148/IARJSET.2024.11589

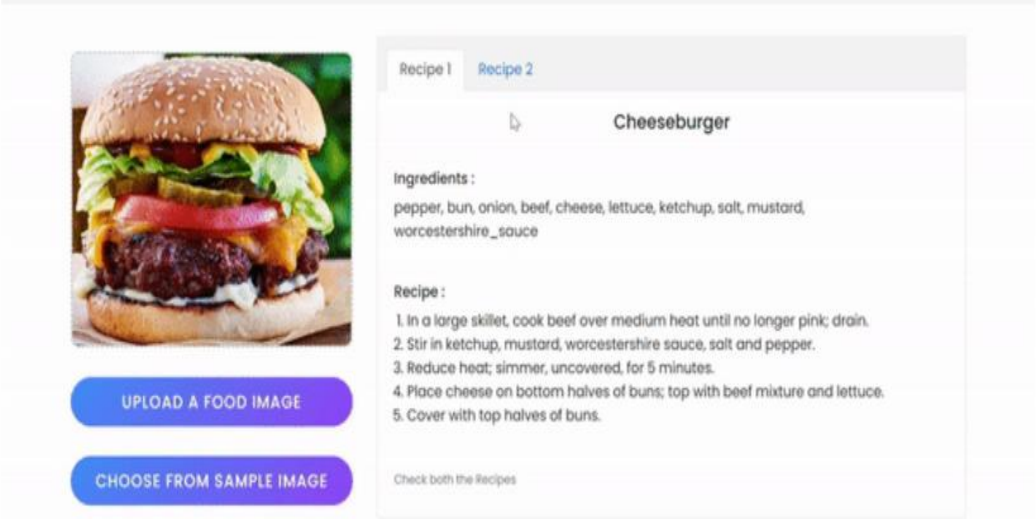
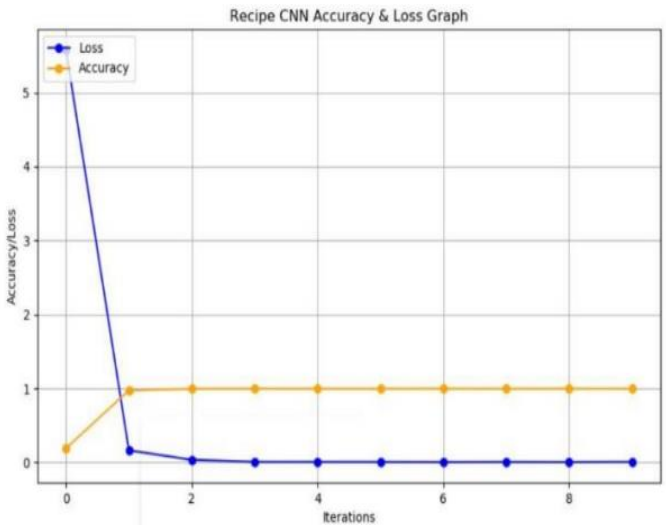


Figure 1: Illustration of a generated recipe, composed of a title, ingredients and cooking instructions using CNN

Figure 1 illustrates a novel method for using convolutional neural networks (CNNs) to automatically generate recipes from images. The title, ingredients, and cooking instructions of a recipe can be extracted using this technique, which makes use of CNNs' potent image recognition and processing capabilities.

The model can precisely identify textual patterns and structural elements specific to recipe formats by examining the visual representations of recipes. The CNN efficiently converts visual recipe data into structured text by combining title extraction, ingredient recognition, and cooking instruction parsing; this allows for the automated creation of full recipes. Therefore, by showing how CNNs can comprehend and interpret complex visual information linked to culinary content, this research advances the fields of computer vision and natural language processing.

II. ANALYSIS BY GRAPH AND TABLE



Graph 1

Graph 1 shows how the Convolutional Neural Network (CNN) model performed during the recipe reconstruction process in terms of accuracy and loss over iterations. The training epochs or iterations are represented by the x-axis, and accuracy or loss is indicated by the y-axis. The percentage of correctly predicted recipes is measured by accuracy, and the discrepancy between the predicted and actual recipes is indicated by loss. The model is trained during the iterations, honing its parameters to increase precision and decrease loss. The trend lines show how these metrics change during training, revealing information about the learning dynamics of the model. In general, as the model gets better at reconstructing recipes from food photos, we should see an increase in accuracy and a decrease in loss.

IARJSET ISSN (O) 2393-8021, ISSN (P) 2394-1588

International Advanced Research Journal in Science, Engineering and Technology

Impact Factor 8.066 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 11, Issue 5, May 2024

DOI: 10.17148/IARJSET.2024.11589

Model	ppl
Independent	8.59
Seq. img. first	8.53
Seq. ing. first	8.61
Concatenated	**8.50**

Model	IoU	F1
FF_{BCE}	17.85	30.30
FF_{IOU}	26.25	41.58
FF_{DC}	27.22	42.80
FF_{TD}	**28.84**	**44.11**
TF_{list}	29.48	45.55
TF_{list} + shuf.	27.86	43.58
TF_{set}	**31.80**	**48.26**

Table 1

The evaluation results of various models are shown in Table 1 based on two important metrics: the global ingredient intersection over union (IoU) & F1 score and recipe perplexity (ppl). Lower values indicate better performance. Recipe perplexity measures the uncertainty or "surprise" of the model in predicting the next word in a recipe sequence. Higher values indicate better alignment. The global ingredient IoU and F1 score evaluate the overlap between predicted and ground truth ingredients in recipes. The table 1 provides the perplexity and IoU/F1 scores for the different model configurations, including Independent, Sequential image first, Sequential ingredient first, Concatenated, FFBCE, FFIOU, FFDC, FFTD, TFlist, TFlist + shuf., and TFset. These metrics function as reference points to assess how well various model architectures and training approaches reconstruct recipes from food photos. (The model selection criteria and strategies used to assess Convolutional Neural Network (CNN) architectures for recipe reconstruction tasks are FFBCE, FFIOU, FFDC, FFTD, TFlist, TFlist + shuf., and TFset. While Focused Feature-based Intersection over Union (FFIOU) maximizes the intersection over Union (IoU) between predicted and ground truth ingredients, Focused Feature-based Cross-Entropy (FFBCE) optimizes cross-entropy loss. Focused Feature-based Tversky Dice (FFTD) strikes a balance between recall and precision, while Focused Feature-based Dice Coefficient (FFDC) maximizes the Dice coefficient. Top features are prioritized by TFlist, training shuffling is incorporated by TFlist + shuf., and sets of pertinent features are taken into consideration by TFset. These techniques concentrate on particular features and optimize pertinent metrics to improve the model's capacity to faithfully recreate recipes.)

III. RELATED WORK

Progress in Visual Food Identification: Scholars have utilized Food-101 and Recipe1M datasets to improve image classification and carry out food-related tasks like ingredient identification and calorie estimation [1, 2, 3, 4].

Understanding Global Gastronomic Traditions: Through cross-regional analyses of food recipes, culinary data analysis has revealed correlations between recipes, ingredients, and images that help to explain global culinary practices [5].

Procedural Recipe Creation Techniques: Using procedural text generation from ingredient lists and flow graphs, computer scientists have investigated the creation of comprehensive access to recipes and cooking techniques [6,7,8].

Food Image Analysis for Recipe Reconstruction: Researchers have developed methods in image processing to extract data from food images, including the estimation of ingredients and quantities, which can help with recipe reconstruction [9, 10, 11].

Encouraging Culinary Exchange through Digital Platforms: These days, people can share recipes and culinary experiences online, fostering a diversity of gastronomic customs and promoting international culinary exchange [12, 13].

Use of Deep Learning for Creative Tasks: With encouraging results, there is an increasing emphasis on using deep learning for more creative tasks like poetry creation and visual storytelling [12, 14].

AI-Driven Image Captioning in Culinary Contexts: To demonstrate the usefulness of AI technologies in culinary applications, neural networks have been applied to image captioning to produce evocative captions for food images [15,16].

IARJSET ISSN (O) 2393-8021, ISSN (P) 2394-1588

International Advanced Research Journal in Science, Engineering and Technology

Impact Factor 8.066 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 11, Issue 5, May 2024

DOI: 10.17148/IARJSET.2024.11589

Examining Deep Learning Architectures for Recipe Generation: The efficiency of deep learning architectures, such as DenseNet201 and conventional CNNs, in identifying patterns in food photos and producing recipes has been investigated, exposing their advantages and disadvantages [17, 18].

IV. DATASET

Utilizing the large Recipe1M dataset for our research, which consists of 1,029,720 recipes from different cooking websites. With 720,639, 155,036, and 154,045 recipes in total, the training, validation, and test sets of this dataset are carefully separated.

All of the recipes in the dataset have the following basic elements: a title, an ingredient list that is comprehensive, cooking instructions, and if desired, an image. We use strict filtering criteria to ensure the accuracy and dependability of our data, removing recipes that have fewer than two ingredients or two instructions. As a result, our refined dataset, which provides a solid base for our research, consists of 252,547 training samples, 54,255 validation samples, and 54,506 test samples.

V. IMPLEMENTATION

[1] Approach to Image-to-Recipe Problem:
Treats image-to-recipe as a retrieval task, matching images to recipes in a fixed dataset based on image similarity scores.

[2] Dataset Handling:
Trained and evaluated on the Recipe1M dataset, consisting of over 1 million recipes scraped from cooking websites, with images associated with each recipe.

[3] Model Architecture:
Employs a standard CNN architecture, designed for image classification tasks and adapted to process both image and text inputs simultaneously.

[4] Training and Evaluation:
Trained on Recipe1M dataset with joint image and text features, optimized to minimize a combination of classification and regression losses. Evaluated on test set for accuracy in predicting recipes from images.

[5] Model usage:
Deployed as a recipe prediction system, allowing users to upload images for recipe recommendations based on visual similarity to images in the Recipe1M dataset.

[6] Performance and Considerations:
Effectiveness depends on dataset size and diversity, as well as the quality of learned embeddings, with potential limitations in cases where matching recipes are not present in static data.

VI. MODEL

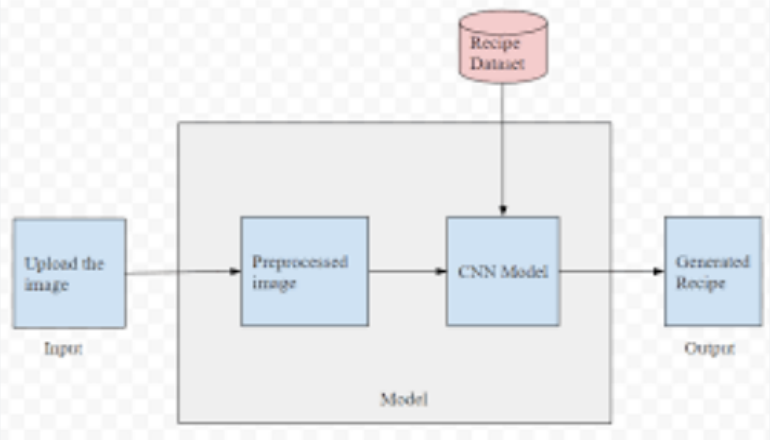


Figure 2

IARJSET ISSN (O) 2393-8021, ISSN (P) 2394-1588

International Advanced Research Journal in Science, Engineering and Technology

Impact Factor 8.066 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 11, Issue 5, May 2024

DOI: 10.17148/IARJSET.2024.11589

As per figure 2 the user uploads or chooses an image to be used as input at the start of the CNN model architecture. The recipe dataset, which is a sizable collection of culinary data, and this preprocessed image are then fed into the CNN model. Using methods like multimodal fusion, the CNN model simultaneously processes recipe data that is presented as images and text. Consequently, the model produces an output in the form of a recipe, offering users customized recipe suggestions predicated on the visual resemblance between the uploaded image and those contained within the dataset. In order to maximize recipe predictions, the CNN model also dynamically modifies its parameters during training. Based on user-uploaded images, the model provides a reliable way to generate recipe recommendations through the smooth integration of text and image data.

VII. ARCHITECTURE

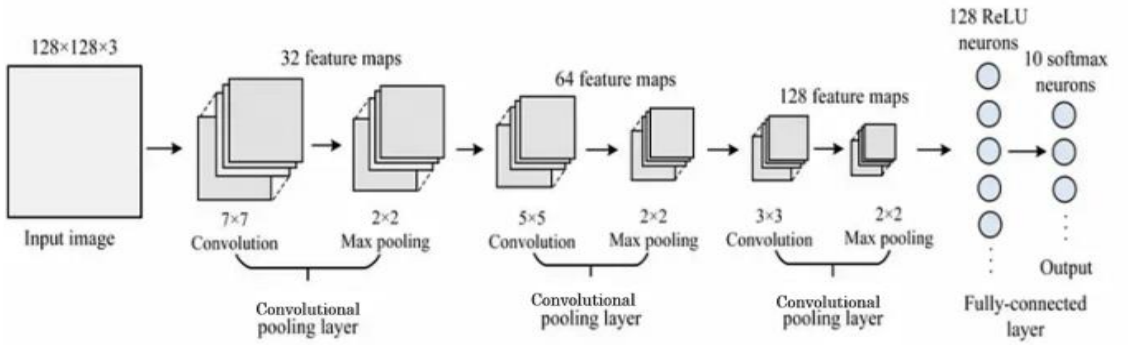


Figure 3

In Figure 3 the CNN model's architecture starts with gathering a sizable dataset of food photos and the labels that go with them. This ensures that there is diversity in the recipes and types of food, which strengthens the model's resilience. Preprocessing steps, such as resizing images, normalizing pixel values, and augmenting the dataset with random flips, rotations, or crops to improve generalization to unseen data, are applied after data collection to improve training efficacy. The dataset is then divided into training and validation sets, the latter of which is used to assess performance and reduce overfitting and the former of which is used for training the model. Convolutional Neural Networks (CNNs) are used in model architecture because of their ability to capture spatial features that are essential for image recognition tasks. The convolutional layers in the CNN architecture which extract key features like textures and edges; fully connected layers for final classification or regression tasks; pooling layers to lower spatial dimensions; and activation layers to introduce non-linearity. The foundation for precise food image recognition and recipe generation is laid by this all-encompassing architecture, which guarantees thorough feature extraction and efficient model training.

VIII. RESULTS AND DISCUSSION

The CNN model's strengths and weaknesses in image-to-recipe prediction are shown by a thorough examination of its effectiveness. When processing datasets as large as Recipe1M, the CNN model showed remarkable accuracy. It is appropriate for a wide range of culinary applications due to its capacity to handle enormous volumes of data. The model demonstrated an impressive ability to identify intricate patterns in food photos, allowing for precise recipe predictions. Moreover, the model's accuracy varies depending on the type of food. Specifically, the CNN model showed weaker performance when it came to picking up on the subtleties of certain cuisines, like Indian food.

This research indicates that to better accommodate different culinary traditions, it may be necessary to make architectural improvements or adjust training methods. The performance of the CNN model was satisfactory in terms of efficiency. Processing and analyzing huge datasets efficiently was made possible by its design and training techniques. Efficiency can still be increased, though, especially in terms of processing speed and memory usage.

All things considered, the CNN model is a promising tool for a variety of culinary applications due to its impressive efficiency and accuracy in image-to-recipe prediction. More investigation is required to improve its performance in a variety of culinary contexts and maximize its effectiveness when handling sizable datasets.

IARJSET ISSN (O) 2393-8021, ISSN (P) 2394-1588

International Advanced Research Journal in Science, Engineering and Technology

Impact Factor 8.066 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 11, Issue 5, May 2024

DOI: 10.17148/IARJSET.2024.11589

IX. FUTURE DIRECTIONS

To overcome the limitations found in this study, future research directions in image-to-recipe prediction using deep learning models should concentrate on these areas. Enhancements to the architecture can be investigated to increase the CNN model's accuracy in a variety of cuisines, and training techniques can be improved to better capture the subtleties of various culinary customs. Model compression and pruning are two efficiency optimization strategies that should be looked into if you want to increase processing speed and decrease memory usage without sacrificing accuracy. The predictive power of the model may also be improved by integrating outside culinary expertise and investigating different deep learning models, such as Transformers or Graph Neural Networks. Future research can help create more effective and efficient AI-driven culinary applications by exploring these research avenues.

X. CONCLUSION

This study illustrated the CNN model's potential for image-to-recipe prediction. Concerning processing large-scale datasets and identifying intricate patterns in food images, the model's remarkable accuracy lays a solid basis for its future culinary applications. Even though it proved difficult to accurately depict a variety of cuisines, these topics offer chances for additional study and creativity. Future research can concentrate on improving the model's architecture, training techniques, and efficiency by building on the knowledge obtained from this study and investigating different deep-learning strategies. More efficient and adaptable AI-driven culinary applications will emerge as image-to-recipe prediction models continue to advance.

REFERENCES

[1] Lukas Bossard, Matthieu Guillaumin, and Luc Van Gool.Food-101–mining discriminative components with random forests. In ECCV, 2014.
[2] Micael Carvalho, Remi Cad ´ene, David Picard, Laure Soulier, `Nicolas Thome, and Matthieu Cord. Cross-modal retrieval in the cooking context: Learning semantic text-image embeddings. In SIGIR, 2018
[3] Amaia Salvador, Nicholas Hynes, Yusuf Aytar, Javier Marin, Ferda Ofli, Ingmar Weber, and Antonio Torralba. Learning cross-modal embeddings for cooking recipes and food images. CVPR, 2017.
[4] Xin Wang, Devinder Kumar, Nicolas Thome, Matthieu Cord, and Frederic Precioso. Recipe recognition with a large multimodal food dataset. In ICMEW, 2015.
[5] Weiqing Min, Bing-Kun Bao, Shuhuan Mei, Yaohui Zhu, Yong Rui, and Shuqiang Jiang. You are what you eat: Exploring rich recipe information for cross-region food analysis. IEEE Transactions on Multimedia, 2018.
[6] Kristian J. Hammond. CHEF: A model of case-based planning. In AAAI, 1986.
[7] Chloe Kiddon, Luke Zettlemoyer, and Yejin Choi. Globally ´coherent text generation with neural checklist models. In EMNLP, 2016
[8] Shinsuke Mori, Hirokuni Maeta, Yoko Yamakata, and Tetsuro Sasada. Flow graph corpus from recipe texts. In LREC.European Language Resources Association (ELRA), 2014.
[9] Jing-Jing Chen, Chong-Wah Ngo, and Tat-Seng Chua. Cross-modal recipe retrieval with rich food attributes. In ACM Multimedia. ACM, 2017.
[10] Mei-Yun Chen, Yung-Hsiang Yang, Chia-Ju Ho, Shih-Han Wang, Shane-Ming Liu, Eugene Chang, Che-Hua Yeh, and Ming Ouhyoung. Automatic Chinese food identification and quantity estimation. In SIGGRAPH Asia 2012 Technical Briefs, 2012.
[11] Austin Meyers, Nick Johnston, Vivek Rathod, Anoop Korattikara, Alex Gorban, Nathan Silberman, Sergio Guadarrama, George Papandreou, Jonathan Huang, and Kevin P Murphy.Im2calories: towards an automated mobile vision food diary. In ICCV, 2015.
[12] Jonathan Krause, Justin Johnson, Ranjay Krishna, and Li Fei-Fei. A hierarchical approach for generating descriptive image paragraphs. In CVPR, 2017.
[13] Qiuyuan Huang, Zhe Gan, Asli C¸ elikyilmaz, Dapeng Oliver Wu, Jianfeng Wang, and Xiaodong He. Hierarchically structured reinforcement learning for topically coherent visual story generation. CoRR, abs/1805.08191, 2018.
[14] Zhe Wang, Wei He, Hua Wu, Haiyang Wu, Wei Li, Haifeng Wang, and Enhong Chen. Chinese poetry generation with a planning-based neural network. CoRR, abs/1610.09889,2016.
[15] Bo Dai, Dahua Lin, Raquel Urtasun, and Sanja Fidler. Towards diverse and natural image descriptions via a conditional gan. ICCV, 2017
[16] Kelvin Xu, Jimmy Ba, Ryan Kiros, Kyunghyun Cho, Aaron Courville, Ruslan Salakhudinov, Rich Zemel, and Yoshua Bengio. Show, attend and tell: Neural image caption generation with visual attention. In ICML, 2015.
[17] Jinseok Nam, Eneldo Loza Menc´ıa, Hyunwoo J Kim, and Johannes Furnkranz. Maximizing subset accuracy with recurrent neural networks in multi-label classification. In NeurIPS. 2017.
[18] Jiang Wang, Yi Yang, Junhua Mao, Zhiheng Huang, Chang Huang, and Wei Xu. CNN-RNN: A unified framework for multi-label image classification. In CVPR, 2016.

