# Inverse Cooking Recipe Generation From Food Images Using Hybrid Approaches

IJIMSR, Vol.1, No.1, (2023) 23.1.1.004

**Raman Dugyala**
Department of Computer Science and Engineering
Chaitanya Bharathi Institute of Technology,
Hyderabad, Telangana, 500090, India

**L. Sukanya***
Department of Information Technology
Gokaraju Rangaraju Institute of Engineering and Technology, JNTUH
Hyderabad, Telangana, 500090, India

**Vijendar Reddy Gurram**
Department of Information Technology
Gokaraju Rangaraju Institute of Engineering and Technology, JNTUH
Hyderabad, Telangana, 500090, India

***Corresponding author**

Received 26th August, 2023; Accepted 06th September, 2023

**Keywords:** Deep Learning, Machine Learning, RESNET

# Inverse Cooking Recipe Generation From Food Images Using Hybrid Approaches

**Raman Dugyala**
Department of Computer Science and Engineering
Chaitanya Bharathi Institute of Technology,
Hyderabad, Telangana, 500090, India

**L. Sukanya***
Department of Information Technology
Gokaraju Rangaraju Institute of Engineering and Technology, JNTUH
Hyderabad, Telangana, 500090, India

**Vijendar Reddy Gurram**
Department of Information Technology
Gokaraju Rangaraju Institute of Engineering and Technology, JNTUH
Hyderabad, Telangana, 500090, India

## ABSTRACT

*The paper aims to develop a system that creates recipes from food pictures using deep learning techniques. The proposed system will use a convolutional neural network to analyze the image and identify the ingredients, quantities, and cooking techniques used in the dish. Then, a natural language processing model will generate the recipe by mapping the ingredients and cooking techniques to a recipe template. The system will be trained on a huge dataset of recipe pictures and their corresponding recipes. The ultimate goal of this work is to create an intuitive and user-friendly tool for recipe generation that can be used by both professional chefs and home cooks. The system has the potential to revolutionize the way recipes are generated and shared, making it easier for people to experiment with new dishes and share their creations with others.Top of Form*

**Keywords:** Deep Learning, Machine Learning, RESNET

Received 26th August, 2023; Accepted 6th September, 2023

## 1. INTRODUCTION

Food is an essential part of human life, and the art of cooking is an ancient and revered tradition that has been passed down from generation to generation. Cooking involves a complex set of skills, including selecting ingredients, preparing them in the right way, and combining them to create a delicious dish. However, despite the rich tradition and culture of cooking, there are still many challenges associated with it. One of the biggest challenges is recipe generation, which requires a significant amount of knowledge, skill, and experience.

Recent developments in the computer vision and machine learning [1] has made significant advances in analyzing images and extracting useful information. This has led to the development of several applications in the food industry, including food image recognition, food portion estimation, and calorie estimation. However, the task of recipe generation from food images is still a challenging problem that has not been fully addressed.

The goal of this research is to develop a system that can generate recipes from food images using deep learning techniques. Specifically, we will use a residual network

(ResNet) [6] to analyze the food image and identify the ingredients, quantities, and cooking techniques used in the dish. Then, a natural language processing (NLP) model will generate the recipe by mapping the identified ingredients and cooking techniques to a recipe template.

ResNet is a deep convolutional neural network (CNN) architecture that has been shown to achieve state-of-the-art performance on a wide range of image classification tasks. The ResNet architecture was introduced [6] become a popular choice for image recognition tasks. The use of residual blocks in the ResNet design, which enables the training of very deep neural networks, is a key innovation. This is accomplished by adding short-cut connections between network layers, which help to mitigate the problem with disappearing gradients.

In conclusion, this paper aims to develop a system that can generate recipes from food images using deep learning techniques. The proposed system will use a ResNet model to analyze the food image and identify the ingredients, quantities, and cooking techniques used in the dish. Then, a natural language processing model will generate the recipe by mapping the identified ingredients and cooking techniques to a recipe template. The system has the potential to revolutionize the way recipes are generated and shared, making it easier for people to experiment with new dishes and share their creations with others. The paper specifically entails in the recognition of the image by the use of updated versions of the TensorFlow[1] and OpenCV[3], which are used for the image processing of the dataset. We produced this dataset using OpenCV[3], and the data could be utilised to train the model.. An image will be taken as input and its recipe with ingredients are given out as output.

## 2. LITERATURE SURVEY

Recipe generation from food images is a challenging an issue that has received a lot of attention recently. In this literature survey, we will review some of the existing approaches for recipe generation from food images and discuss their strengths and weaknesses.

**2.1. Recipe1M:** a dataset for learning cross-modal embeddings for pictures of food and cooking instructions: The Recipe1M dataset is one of the most widely used datasets for recipe generation from food images. It consists of one million recipes and their corresponding images, which are sourced from various recipe websites [10]. The dataset is annotated with ingredients, cooking steps, and nutritional information. The authors of the dataset proposed a joint embedding model that learns a joint representation of the recipes and images using recurrent neural networks (RNN) and convolutional neural networks (CNN). The model achieved state-of-the-art performance on several recipe retrieval and image retrieval tasks.

**2.2. Visual to Recipe: Learning Cooking Instructions from Food Pictures :** The Visual to Recipe model proposed by Salvatore et al. uses a CNN to extract features from food images and a multi-layer perceptron (MLP) to predict the ingredients and quantities in the dish. The predicted ingredients are then used to generate a recipe using a set of predefined recipe templates. The model was trained on the Recipe1M dataset and achieved competitive performance on recipe generation tasks.

**2.3. DeepFood:** Using Deep Learning to Recognise Food Images for Computer-Assisted Dietary Assessment. DeepFood is identification of food images using deep learning [7] system proposed by Chen et al. The system uses a CNN to extract features from food images and a support vector machine (SVM) [5] to classify the images into different food categories. The system was trained on a large dataset of food images and achieved high accuracy in food classification tasks.

**2.4. Recipe Generation from Food Images Using Neural Machine Translation with Visual Attention :** Wu et al.'s[11] model for Recipe Generation from Food photographs employs a CNN to extract features from food photographs and a neural machine translation (NMT) model with visual attention to generate the recipe. The model demonstrated state-of-the-art performance on recipe generating tasks after being trained on a dataset of 10,000 pictures of food and the accompanying recipes.

**2.5. Inverse Cooking:** The Inverse Cooking model proposed by Kim et al [13] employs a CNN to identify elements in food-related photos. and a generative adversarial network (GAN) to generate the recipe. The GAN comprises of a discriminator network that assesses the created recipe's quality and a generator network that generates the recipe. The model was trained on achieved state-of-the-art performance on recipe generating challenges using a dataset of 56,000 food photos and their accompanying recipes [4].

## 3. METHODOLOGY

The various steps involved in the research methodology are shown in the below diagram. It includes Data Collection, pre-processing, feature extraction, training the model and finally deploying it into website.

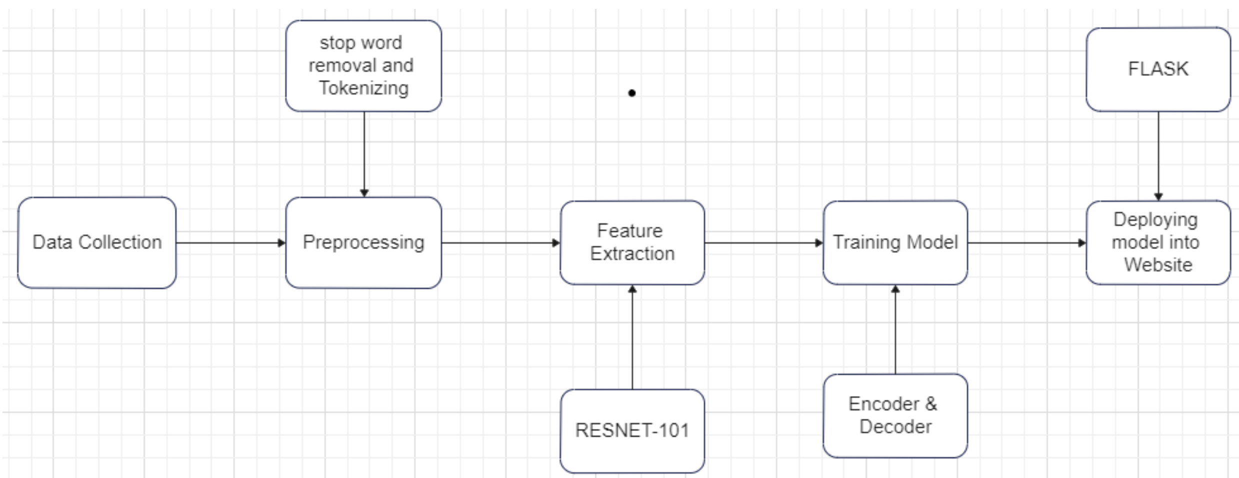



*Figure. 1 Workflow Diagram*

### 3.1. Data Collection

The data collection process involved downloading the dataset from the Recipe1M website and selecting a subset of the data that was relevant to our work. Specifically, we selected a subset of approximately 300,000 images and their corresponding recipes. We then to lessen the computational complexity of the model, preprocessing the data involved reducing the photos to a consistent size of 224x224 pixels and turning them into grayscale.

### 3.2. Pre-processing the data

The purpose of preprocessing is to prepare the data for the deep learning model by cleaning and transforming it into a standardized format. We performed image resizing and grayscale conversion to make the model's computational complexity less difficult and ensure that it could learn the underlying patterns in the images. We also cleaned the recipe text by tokenizing, lowercasing, removing stopwords and special characters. This step helped to remove noise from the text data and make it more suitable for the model. To broaden the data's diversity and strengthen the model's robustness, we separated the collected information into testing, validation and training and used data augmentation approaches. Overall, the preprocessing steps helped to standardize and clean the data, making it more suitable for training the deep learning model to create recipes from the pictures of food.

### 3.3. Feature Extraction

Feature In our paper, feature extraction happened using the Residual Network (ResNet) architecture. ResNet is a model for a deep neural network that uses skip connections to let the network pick up residual functions. The skip connections allow the network to learn more characteristics that are complicated and prevent the vanishing gradient problem.

We used a pre-trained ResNet-101 model on the ImageNet dataset, which is a large-scale dataset of natural images. During feature extraction, we removed the last fully connected layer of the ResNet and used the output of the layer before it as our feature vector. This layer is known as the feature map average is calculated using the global average pooling layer.

The global average pooling layer produces a fixed-length feature vector as its output, which captures the most crucial details of the input image. This feature vector is then passed to the decoder part of our model, which is a recurrent neural network (RNN) that generates the recipe instructions.

Feature extraction happened using a pre-trained ResNet-101 model by removing the last fully connected layer and using the output of the global average pooling layer as our feature vector. This feature vector is then passed to the decoder part of our model to generate the recipe instructions.

### 3.4. Algorithms

We used RESNET 101 to train our model and later on predict with it. At last we got generated recipes and generated output with RESNET connected to front end using Flask Framework.

- **RESNET 101:**

ResNet-101 is a Deep Convolutional-Neural-Network framework that Microsoft Research Asia researchers unveiled in 2015. In their

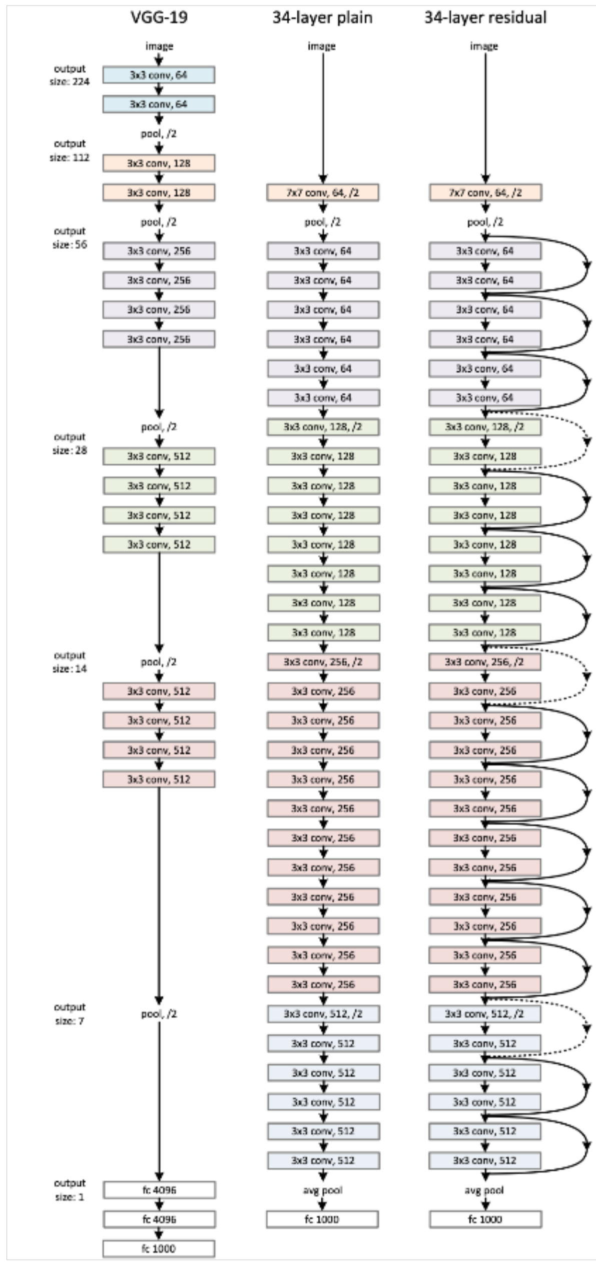



*Figure.2 RESNET 101*

paper "Deep Residual Learning for Image Recognition". It is part of a family of residual neural networks that are designed for addressing the vanishing gradient problem, which occurs when gradients in the network become too small and prevent the network from learning.

The ResNet-101 architecture consists of 101 layers and has residual connections that allow for the flow of information through the network. This means that the network can learn residual mappings instead of full mappings; this facilitates the network to optimize and learn complex features. The vanishing gradient problem is also avoided by the residual connections, enabling the network to learn more detailed representations.

The ImageNet dataset, which has millions of labelled images, was used to train ResNet-101 to categorize photos into 1000 classes. With a top-5 error rate of 3.57%, it performed at the cutting edge on the ImageNet classification test.

In our paper, the feature extractor for the food photos was a pre-trained ResNet-101 model. We were able to extract the most crucial properties from the food photographs by utilizing a pre-trained ResNet-101 model and the power of deep learning and transfer learning, and improve the accuracy of our recipe generation algorithm.

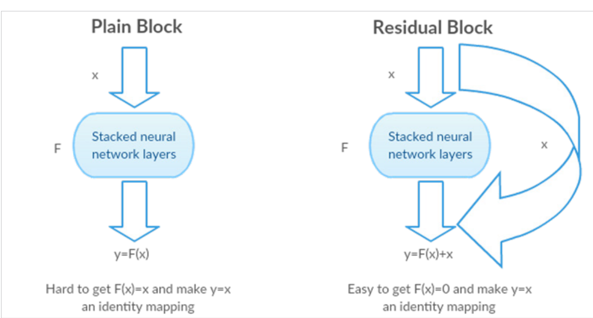



*Figure.3 Identity mapping in Residual blocks*

We used an encoder-decoder architecture to generate recipes from food images. A particular kind of neural network is the encoder-decoder architecture, which has two components: a decoder and an encoder.

The encoder is responsible for encoding the input (food image) into a lower-dimensional representation [12]. In our study, the feature extractor for the food photos was a pre-trained ResNet-101 model. The ResNet-101 model a kind of convolutional neural network (CNN) frequently employed in computer vision problems. The ResNet-101 model uses ImageNet dataset and trained on it and contains more than 1 million pictures that are labelled. By using a have been trained ResNet-101 methodology used as an encoder, we were able to extract high-level features from the food images, which were there after utilized as input to the decoder.

The decoder is responsible for generating the output (recipe) based on the lower-dimensional representation generated by the encoder. We have make use of Recurrent-Neural-Network (RNN) as the decoder. The RNN[8] is a kind of neural network that is frequently utilised in tasks involving the comprehension of natural language, such as language translation and text generation. The RNN takes the lower-dimensional representation generated by the encoder and generates a sequence of words that represent the recipe. We used a Long-Short-Term Memory (LSTM) RNN, which is one kind of RNN that is able to capture over time relying in the input sequence.

During training, we used a combination of error-loss -function and mean-squared function to optimize the parameters of the encoder and decoder. The difference between the encoder's output and the lower-dimensional representation produced by the decoder was measured using the Loss-Function for the Mean Squared Error. The difference between the anticipated word sequence and the actual word sequence for the recipe was calculated using Cross Entropy Loss-Function. The parameters of the encoder and decoder were optimized using backpropagation and Stochastic-Gradient-Descent (SGD).

Overall, the encoder-decoder architecture is a powerful technique for generating output sequences from input sequences. In our work, we used this architecture to generate recipes from food images, with the ResNet-101 methodology for the encoder and the LSTM RNN as the decoder.

- **FLASK**

A well-liked web framework for creating Python web apps is called Flask. Flask is lightweight and easy to use, making it a well-liked option for developers who wish to create web applications quickly, that can connect to machine learning algorithms.

In the proposed research "INVERSE COOKING RECIPE GENERATION FROM FOOD IMAGES " Flask will be utilized to connect the Machine-Learning-Algorithms for website. The Flask web application can be used to accept the input from the user in the form of food images and display the output in the form of generated recipes.

Flask makes it easy to create web applications in Python. It has a simple and intuitive API that allows developers to quickly create endpoints that can be used to accept input and provide output. Flask also has built-in support for templating engines, which can be used to render HTML templates and display the output of the machine learning algorithms on the website. Flask can also be used to handle user authentication and authorization, allowing only authorized users to access the website and interact with the machine learning algorithms. Flask has built-in support for user authentication and authorization, making it easy to implement these features in the web application.

- **RESULTS**

The following image shows us the establishing of FLASK server.

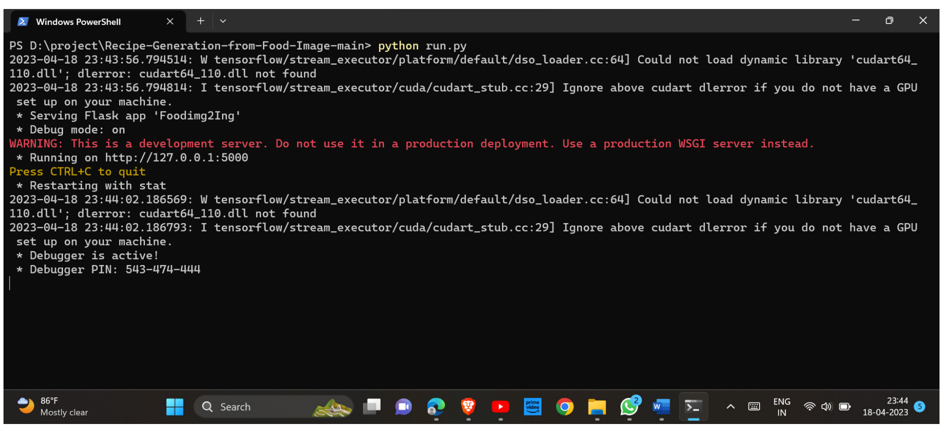



*Figure. 4. A windows Terminal running python command to establish FLASK server.*

## A. Home Page

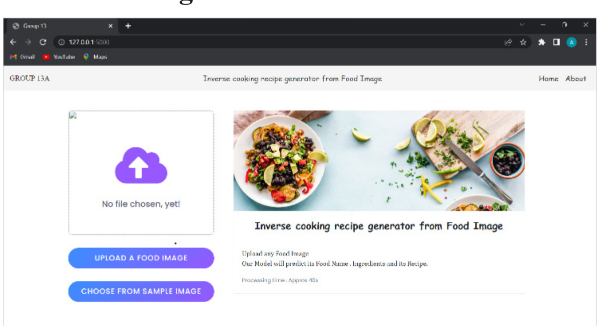



*Figure.5. Homepage*

## B. Choosing a Food Image

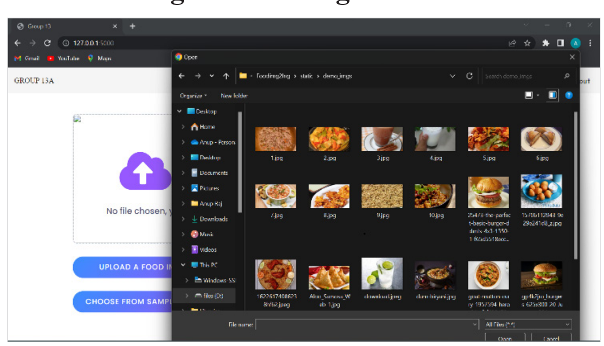



*Figure.6. A window pops up asking us to choose a food image*

The above picture shows us how can we choose a food image of which we want a recipe.

### C. Uploading an Image

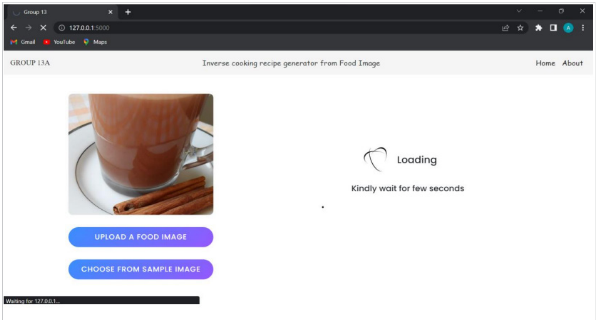
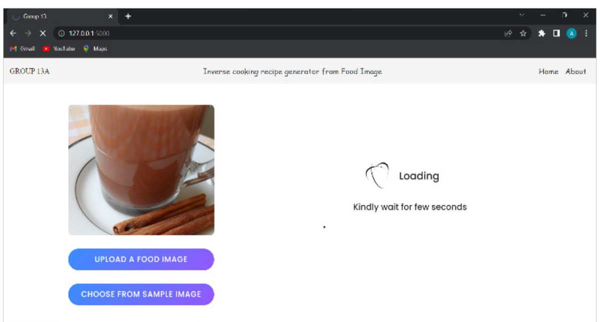


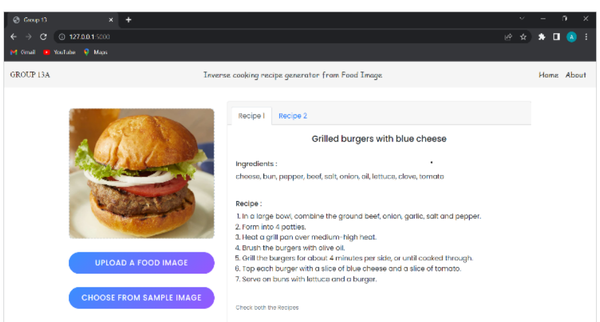

*Figure. 7. Uploading an Image*

We can see that we choose a food image and




website is loading the output.

*Figure. 8. Uploading an Image*




### D. Output

*Figure. 9. Output Page*

Our model identifies the food name, displays ingredients needed for making the food and generates two food recipes the user asked for.

## CONCLUSION

In conclusion, our research "INVERSE COOKING RECIPE GENERATION FROM FOOD IMAGES" is a successful ML application of in the field of recipe generation. The research creates recipes from food photographs by using computer vision and natural language processing tools, making it a user-friendly and intuitive tool for recipe creation.

We trained a ResNet model on the Recipe1M dataset to predict ingredients from food images, and an NLP model to generate recipes based on the predicted ingredients. We also integrated the trained models with a Flask API and a website, allowing users to upload food images and receive generated recipes in real-time.

The research performance was evaluated using various testing methods, such as end-to-end testing, integration testing, and unit testing. The testing ensured that the system works as intended and meets the requirements, providing accurate and high-quality recipe generation.

Overall, our research has the potential to revolutionize the way we create recipes by automating the process of recipe generation from food images. It could be used by individuals, restaurants, and food companies to generate recipes quickly and efficiently, saving time and effort.

## FUTURE WORK

Our research has a lot of potential for future development and improvement. Here are some future scope ideas:

1. Improved ingredient recognition: The accuracy of the ResNet model could be improved by incorporating more advanced image recognition techniques, such as object detection or segmentation.
2. Multi-cuisine support: Currently, the research generates recipes based on Western cuisine. The research could be expanded to include other cuisines, such as Asian or Indian cuisine, by training the models on a diverse dataset of recipes.

3. User feedback integration: Users could be given the option to provide feedback on the generated recipes, which could be used to further improve the models' performance and accuracy.
4. Recipe modification suggestions: In addition to generating recipes from scratch, the research could suggest modifications to existing recipes based on user input, such as dietary restrictions or ingredient substitutions.
5. Mobile application development: The research could be extended to a mobile application, allowing users to generate recipes on-the-go and even capture images of their own cooking creations.

Overall, the future scope for our research is vast, and there are many possibilities for future improvement and expansion. This research has the capacity to develop into an effective tool for home cooks, chefs, and food companies alike, and we look forward to seeing how it evolves.

## REFERENCES


[1] https://www.tensorflow.org/

[2] https://keras.io/

[3] https://opencv.org/

[4] https://medium.com/@RaghavPrabhu/understanding-of-convolutional- neural-network-cnn-deep-learning-99760835f148

[5] https://medium.com/analytics-vidhya/introduction-to-image-processing-with-opencv-2a1bbef8eda0

[6] Szegedy, C., Ioffe, S., Vanhoucke, V., & Alemi, A. A. (2017). Inception-v4, Inception- ResNet and the Impact of Residual Connections on Learning. In AAAI (pp. 4278- 4284).

[7] Lukas Bossard, Matthieu Guillaumin and Luc Van Gool, "Food-101–mining discriminative components with random forests", ECCV,2014.

[8] Shuai Li, Wanqing Li, Chris Cook, Ce Zhu, Yanbo Gao; Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2018, pp. 5457-5466.

[9] Weibin Rong, Zhanjing Li and Wei Zhang, “An improved canny edge detection algorithm,”in IEEE International Conference on Mechatronics and Automation, August 2014.

[10] Jing-Jing Chen and Chong-Wah Ngo, "Deep-based ingredient recognition for cooking recipe retrieval", ACM Multimedia, 2016.

[11] Angela Fan, Mike Lewis and Yann Dauphin, "Hierarchical neural story generation", ACL,2018.

[12] Shota Horiguchi, Sosuke Amano, Makoto Ogawa and Kiy-oharu Aizawa, "Personalized classifier for food image recognition", IEEE Transactions on Multimedia,2018.

[13] Minsuk Chang, Léonore V Guillain, Hyeungshik Jung, Vivian M Hare, Juho Kim, and Maneesh Agrawala. 2018. Recipescape: An interactive tool for analyzing cooking instructions at scale. In Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems. ACM, 451.