International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 05 Issue: 06 | June - 2021 ISSN: 2582-3930

Prediction of Recipes from Food Images Using Image Processing and Machine Learning

Prof. A. V. Kanade[1] (Under the Guidance),

Khushbu Mahajan[2], Mansi Ekkaldevi[3], Shraddha Shete[4], Priyanka Sirsula[5]

Professor, Department of Information Technology, BVCOEW, Pune, Maharashtra, India
Final Year Students, Department of Information Technology, BVCOEW, Pune, Maharashtra, India

---

Abstract - We all are using social medias like Instagram, Facebook, etc. where we all post too many images of food dishes. But by simply looking at food image we don't have exact idea of cooking instruction. Therefore, in this paper we introduce a cooking system that generates cooking recipes from given food images. The aim of this project is to extract important data from food images and to build a food recommendation system for ingredients and recipes. The system mostly focuses on the predicting the food dish, this estimation of the food is highly challenging task. In this System, we present the recipes corresponding to the given food image and translate the fine-grained ingredient based representation of the food items.

Key Words: Image Processing, Machine Learning, CNN Algorithm, Feature Extraction, Classification.

1. INTRODUCTION

Food has a profound impact on human life and it is fundamental to the human experience. Nowadays, the fast development of various networks, such as social networks and mobile networks allows users to easily share food images, recipes, cooking videos or record food diary via these networks, leading to large-scale food dataset.

Food recognition challenges current computer vision systems to go beyond the merely visible. When compared with natural image understanding, visual ingredient prediction requires high-level processing and prior knowledge. Recipes are the universal language of cooking. Through a recipe you quickly understand how a certain dish can be made. When people are searching for recipes it is now quite common to turn to websites and apps to find food-related content. One very simple application could be for recipe search and retrieval using a picture of a dish, such as something you tried while traveling but whose name you has now forgotten.

Nowadays, Convolutional neural network (CNN) and its derivative algorithms have been recognized as the key methods in most of the surveyed articles , which can automatically learn deep features of input digital information for subsequent classification or regression tasks. The large amount of data collected by the tools for food quality can be successfully processed by CNN. We prefer a concise representation of ingredient, as features of the model, to control the model element sequence. Classification of recipes into respective cuisines based on ingredients has been done by using classification technique. It outputs into sequence of the sentences as per image ingredients respectively. We have performed classification technique and analyzed how they perform in comparison. Model included n number of sample food images in the respective dataset which results into desired output of the recipe generation model.

2. LITERATURE REVIEW

The practice of predicting recipe from food image is done by using LSTM this system takes input as a coloured image. The pre-processing of image is done like clarifying image remove noise from image then LSTM is applied. Long Short-Term Memory (LSTM) networks are a type of recurrent neural network capable of learning order dependence in sequence prediction problems. This is a behaviour required in complex problem domains like machine translation, speech recognition, and more. LSTMs are a complex area of deep learning. It can be hard to get your hands around what LSTMs are, and how terms like bidirectional and sequence-to-sequence relate to the field. It uses two types of LSTM i) Bi-LSTM ii) Sentiment Analysis

Bi-LSTM: (Bi-directional long short term memory): Bidirectional recurrent neural networks(RNN) are really just putting two independent RNNs together. This structure allows the networks to have both backward and forward information about the sequence at every time step. Using bidirectional will run your inputs in two ways, one from past to future and one from future to past and what differs this approach from unidirectional is that in the LSTM that runs backward you preserve information from the future and using the two hidden states combined you are able in any point in time to preserve information from both past and future.

Sentiment Analysis: the process of computationally identifying and categorizing opinions expressed in a piece of text, especially in order to determine whether the writer's attitude towards a particular topic, product, etc. is positive, negative, or neutral.

The respected project chooses LSTM RNN because of its ability to remember sequences of data to generate new data. They modeled a word-based LSTM to consecutively generate words based on the previous words. With the initial sequence of ingredients, the model generates complete recipes, word by word. The model is designed with softmax output ,with two stacked LSTM layers, Each cell unit has a given time step t, to understand the complex nature of the data MIL is used, i.e. ingredients, steps, recipe titles, and categories. The project implemented the new approach to facilitate dynamic interaction with a trained model it shows in figure 3. The MIL allows different types of model output and the additional structure enables fast learning. This uses because of deep learning algorithms require a large amount of training data for the model to perform well. So the Meal-Master is one of the largest free recipe database they used are they filtered out 60k recipes and extracted title, ingredients and instructions of recipe.

Another approach is by using SVM classifier. Support Vector Machine (SVM) is a classification and regression prediction tool that uses machine learning theory to maximize predictive accuracy while automatically avoiding over-fit to the data.

In this system food image is taken in color format and then convert is into grey format for matrix transformation then image pre- processing is applied after that image segmentation is used for extracting information from image as now features are extracted now we use SVM classifier for classifying the image is in which part of dataset it belongs.

The main disadvantage of the SVM algorithm is that it has several key parameters that need to be set correctly to achieve the best classification results for any given problem. Parameters that may result in excellent classification accuracy for problem A may result in poor classification accuracy for problem B. The user may, therefore, have to experiment with a number of different parameter settings in order to achieve a satisfactory result. The SVM does not perform well when the number of features is greater than the number of samples.

SVMs are not very efficient computationally, if your dataset is very big, such as when you have more than one thousand rows. Choosing an appropriate Kernel function (to handle the non-linear data) is not an easy task. It could be tricky and complex. In case of using a high dimension Kernel, you might generate too many support vectors which reduce the training speed drastically.

Over the years, research on convolutional neural networks (CNNs) has progressed rapidly, however the real-world deployment of these models is often limited by computing resources and memory constraints. Convolutional layers are the major building blocks used in convolutional neural network. It is a Deep Learning algorithm which can take in an input image, assign importance i.e. learnable weights and biases to various aspects/objects in the image and be able to differentiate one from the other.

Comparison of CNN over SVM:

1. Convolution Neural Network is non-linear whereas support vector machine is a linear classifier.
2. CNN works well with Visual images recognition whereas SVM is used widely in classification problems.
3. CNN increases model complexity by adding more layers but in SVM increasing model complexity isn't possible.

Hence, the various food recommendation systems helps to generate highly available food instructions corresponding to desired food image and in accounting for the proportion of the various ingredients, with an enhanced number of features, such as adding additional food recommendations, estimation of food features have been proposed in recent years.

3. PROPOSED SYSTEM

In this System, we present the recipes corresponding to the given food image. The system mostly focuses on the predicting the ingredients, this estimation of the food is highly challenging task. The system aim to translate the fine-grained ingredient Based representation of the food items. It predicts the ingredients as a sets or lists and generates cooking instructions by using CNN algorithm and by attending both image encoder and ingredient decoder, & by exploring various attention techniques.

4. IMPLEMENTATION

The system aim to translate the fine-grained ingredient based representation of the food recipe. Image processing can be defined as the technical analysis of an image by using complex algorithms. Most of system takes as input the base recipe name, the sequence of base ingredient then the system provides recipe of food. So, we have proposed a system this system will predict the food recipe from food image by using the algorithm and methods we can

achieve this. Here, food image is used as the input, and by performing various images processing techniques are performed in background in order to get an enhanced image or to extract some useful information from it.

To classify images and label them as different individuals we can do it by analysing the pixel values and looking for patterns in them. It is hard to find patterns by analysing the pixel values alone. Hence we might require a more advanced technique that can detect these edges or find the underlying pattern of different features from the image using which these images can be labelled or classified, here advanced technique like CNN comes into the picture. CNN or the convolutional neural network (CNN) is a class of deep learning neural networks. In short think of CNN as a machine learning algorithm that can take in an input image, assign importance (learnable weights and biases) to various aspects/objects in the image, and be able to differentiate one from the other. CNN works by extracting features from the images. The role of CNN is to reduce the images into a form that is easier to process, without losing features critical towards a good prediction.

CNN accept the image in colour format CNN has 4 steps convolutional layers, pooling layers, flattening and fully-connected layers. We can add more than one convolution layer when building the neural network, where the first Convolution Layer is responsible for capturing gradients whereas the second layer captures the edges. The addition of layers depends on the complexity of the image. We are using 64 bit filters. Image is going through these 4 steps then we will get matrix. We have to fit our image dataset for that we have to pre- process the images to prevent over-fitting, overfitting is when you get a great training accuracy. Prepare the training set as well as the test set of images that are present in proper structured directories, where the directory's name is taking as the label of all the images present in it. For the training, model uses CNN with two different branches, one of which is used for the training of character embedding's , and another that takes as input the explicit representation alongside the pretrained word vectors and keeps them frozen. In the end, the outputs of both branches are concatenated and used for the final classification.

The test_image holds the image that needs to be tested on the CNN. Once we have the test image, we will prepare the image to be sent into the model by converting its resolution to 64x64 as the model only excepts that resolution. Then we are using predict () method on our classifier object to get the prediction. As the prediction will be in a binary form, we will be receiving either 1 or 0, 1 for true if recipe is present in dataset, 0 is for false if recipe is not present in dataset.

5. SYSTEM DESIGN

System design is a process of defining architecture, models, interfaces and data for a system to satisfy specified requirements. System design could be seen as the application of system theory to product developing. There is some overlap with disciplines of system analysis, system architecture and system engineering.

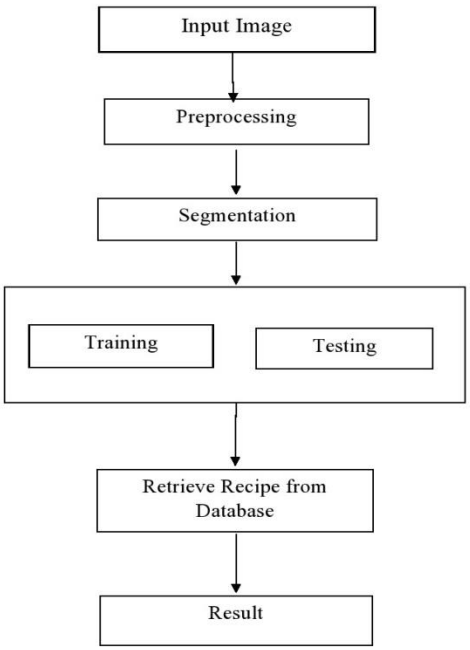

Fig. 1 System Architecture

1. **Input:** First step Give the Input as Food Image.

2. **Preprocessing:** The aim of pre-processing is an improvement of the image data. In preprocessing image convert into RGB to GREY.

3. **Segmentation:** In digital image processing image segmentation is the process of partitioning digital image into multiple segments like sets of pixels, also known as image objects.

4. **Feature extraction:** Feature extraction helps to get the best feature from the input image. In computer vision and image processing, a feature is a piece of information about the content of an image; typically about whether has certain properties. Features may be specific structures in the image such as points, edges or objects.

5. **CNN classification:** Training: Extract the features using CNN and create file of that features.

Testing: In testing the input file get from the training will be compare with the images present in testing images.

6. **Retrieve recipe from database:** According to the label of image the appropriate recipe will fetch the database.

6. RESULT

Our recipe generation system takes a food image as an input and outputs a sequence of cooking instructions. This model generates a cooking recipe which contains a title, ingredients and set instructions directly from an image. These recipes obtained with a method, which firstly predicts the ingredients from an image and then translations on both the image and the ingredients to generate the cooking instructions.

7. CONCLUSION

In this paper, we introduced an image-to-recipe retrieval system, which takes a food image and produces a recipe consisting of a title, ingredients and sequence of cooking instructions. We explored instruction generation conditioned on images. we learnt to create good datasets and build different models to improve upon the accuracy for object recognition also build the dataset affects the model greatly and hence it extremely important to clean the data before using it.

REFERENCES

[1] "Inverse Cooking: Recipe Generation from Food image" Amaia Salvadorl Michal Drozdal2 Xavier Giro-i-Nietol Adriana Romero2 1Universitat Politecnica de Catalunya 2Facebook AI Research, 2019 IEEE

[2] "Food Image to Cooking Instructions Conversion Through Compressed Embedding's using Deep Learning", 2019 IEEE 35th International Conference on Data Engineering Workshop.

[3] "Recipe1M+: A Dataset for Learning Cross-Modal Embeddings for Cooking Recipes and Food Images" Javier Marin, Aritro Biswas, 2019 IEEE.

[4] "Fo rage: Optimizing Food Use With Machine Learning Generated Recipes", Angelica Willis, Elbert Lin, Brian Zhang-2016.

[5] 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," Computer Vision and Pattern Recognition, 2017.

[6] Towards Data Science, "Building a Food Recommendation System," 2020. [Online]. Available: https://towardsdatascience.com/building-a-food-recommendation-system-90788f78691a.

[7] X. Wang, D. Kumar, N. Thome, M. Cord, and F. Precioso. Recipe recognition with large multimodal food dataset. In ICMEW, 2015.

[8] S. Horiguchi, S. Amano, M. Ogawa, and K. Aizawa. Personalized classifier for food image recognition. IEEE Transactions on Multimedia, 2018.