JOURNAL OF ALGEBRAIC STATISTICS
Volume 13, No. 3, 2022, p. 2391 - 2396
https://publishoa.com
ISSN: 1309-3452

Recipes Creation Using Food Images Through Inverse Cooking

Shaik Imam Saheb [1], Ubaidullah Khan [2], Mohd Shahrukh [3], Mohd Imaduddin [4], Amatul Kareem Sariya [5]

[1] Associate Professor, Department of Computer Science and Engineering, Lords Institute of Engineering and Technology, Hyderabad, Telangana, India.
[2, 3, 4, 5] Research Scholar, Department of Computer Science and Engineering,
Lords Institute of Engineering and Technology, Hyderabad, Telangana, India.

Email : *[1] shaikimamsa@gmail.com*

ABSTRACT

People like taking pictures of food because they like food. Behind every meal is a complicated recipe that tells its story. Unfortunately, we can't see how it was made just by looking at a picture of it. So, in this paper, we show an inverse cooking system that can make recipes from pictures of food. Our system guesses ingredients as sets by using a new architecture to model their dependencies without forcing any order. It then makes cooking instructions by looking at both the image and its guesses for ingredients at the same time [8]. We test the whole system on the massive Recipe1M dataset and show that: (1) we improve performance compared to earlier baseline methods for nutritional information prediction; (2) we can get high-quality recipes by using both images and ingredients; and (3) according to human judgement, our system can make more interesting recipes than retrieval-based methods. We share code and models with the public.

Index Terms— An example of a created recipe, which includes a title, ingredients, and instructions for preparation, CNN

I. INTRODUCTION

Humans are unable to survive without food. We are who we are because of it [10, 14], and that is why we use it as a source of energy. Cooking, preparing, eating, and talking about food take up a substantial chunk of our everyday lives as the old adage says. With so many individuals posting images of their meals on social media, food culture has spread like wildfire in the age of the internet [11]. Instagram searches for #food produce at least 300 million results, while searches for #foodie produce at least 100 million results, demonstrating the undeniable importance of food in our culture. Cooking styles have also changed over time, as have eating habits [1]. Food preparation used to be done at home, however nowadays we eat a lot of food cooked by others (e.g. takeaways, catering and restaurants). As a result, precise information regarding cooked food can now be accessed more easily.

It is difficult to determine exactly what we eat because the options are so limited. Thus, we believe that inverse cooking technologies, which can deduce ingredients and cooking directions from a pre-prepared meal, are needed.

Recent years have seen remarkable progress in image recognition tasks such natural picture categorization [7, 14], object recognition [12, 13], and semantic segmentation [7, 9]. Due to its significant intra class variability and substantial deformations that occur during the cooking process, the recognition of food is more difficult than the recognition of natural images. Cooked food often obscures the colours, shapes, and textures of its ingredients. It also involves advanced reasoning and prior information (for example, it is likely that a cake has sugar, but not salt), while a croissant is expected to include butter. As a result, food recognition pushes existing computer image processing algorithms to go beyond the merely observable and combine prior information to provide high-quality structured preparing food descriptions.

Food categorization has been the subject of previous efforts to better comprehend food [1, 9, 14].

It's not enough to recognise the type of meal or its ingredients; a technology for thorough and comprehensive food recognition must also be able to understand the preparation process. Image-to-recipe problem is traditionally phrased as a retrieval challenge where a recipe is found from a fixed dataset using an embedding space similarity score [3, 4, 15]. It is critical that the database size and variety as well as the validity of the learnt embedding be taken into account when analyzing the quality of such systems. When the static dataset does not contain a recipe that matches the picture query, these systems fail.

Image-to-recipe problem can be formulated as conditional generation challenge to avoid dataset limitations of retrieval algorithms. This is why we developed a system that can automatically create a cooking recipe from an image, complete with a title, ingredients, and step-by-step directions. As shown in Figure 1, our algorithm first predicts ingredients from a picture and then conditional on both the image and the components to produce the cooking guidelines. Our technology, to our knowledge, is the only one that generates cooking instructions from photos of food. An image and its expected constituents are used to generate a sequence of instructions in the instruction generation problem. A set prediction problem is used to solve the ingredients estimation problem because of its inherent structure. We don't penalise for prediction order when modelling ingredient dependencies, so the relevance of prediction order is re-examined [5]. We thoroughly test our system upon that large-scale Recipe1M dataset [15], which includes photos, ingredients, and step-by-by-step recipes. Our results are encouraging. Furthermore, in a user assessment report, we demonstrate that our inversion cooking system surpasses previously introduced methods for retrieving recipes using images. Using a small set of photos, we demonstrate that humans struggle with food image-to-ingredient prediction and that our approach is superior to theirs. In this work, we introduce an inverse cooking system that generates cooking instructions based on a picture and its ingredients, investigating various attention strategies to reason about both modalities at the same time [5]. We thoroughly examine components as both a list and as a set, and recommend a new framework for ingredient prediction that takes advantage of founder among ingredients without impositions any assumptions. We illustrate the effectiveness of our proposed system over existing image-to-recipe retrieval systems through a user research that shows ingredient prediction is truly a difficult task.

II. RELATED WORK

It's a huge database of images of Chinese food that may be used to identify it

New and demanding large-scale food image datasets termed "ChineseFoodNet" are introduced in this research, with the goal of automatically detecting depicted Chinese cuisine. For the most part, existing food image collections have been compiled from photographs taken for recipes or from people taking selfies with food. For each food category in our database, we've included images taken not only from online recipes and menus, but also from actual plates, recipes, and menus [6]. There are approximately 180,000 food images on ChineseFoodNet, divided into 208 categories, all of which show a wide range of different ways to portray the same Chinese meal. Data collecting, data cleaning and data labelling, including the use of machine learning algorithms to reduce labour-intensive manual labelling, are all part of our efforts to develop this large-scale image dataset. On Chinese Food Net, we present a comprehensive comparison of several present condition [7] convolutional neural networks (CNNs). As an added bonus, we've developed "Tasty Net," a revolutionary two-step data fusion strategy that combines predictions from various CNNs with a voting method. On the validation set, we get an accuracy of 81.43 percent and an accuracy of 81.55 percent using our proposed approach. ChineseFoodNet's most recent dataset is open to the public and can be seen at https://sites.google.com.

Recipe retrieval based on deep-based ingredient recognition

The calculation of nutrition data made possible by retrieving recipes that correspond to specified dish images is critical for numerous health-related applications. Today's methods mostly focus on the recognition of food categories based on the overall appearance of the dish rather than a detailed investigation of the constituent composition [11]. The problem of zero-shot retrieval is unable to be solved using these methods. Due to the wide range in visual appeal and ingredient composition of food, content-based retrieval without food category knowledge is also challenging to achieve adequate performance. In general, it's more scalable to grasp ingredients underlying recipes than it is to recognise every food category, therefore zero-shot retrieval is appropriate [4]. However, component identification is a much more difficult task than food classification, and thus severely limits the usefulness of depending on them for retrieval.. An approach based on the reciprocal, but also ambiguous, relationships between constituent identification and food categorization is proposed in this work. Recipes may be found in a fraction of a second using feature representations and semantic labels learned from ingredients. Cooking recipe retrieval is a unique zero-shot problem, and this article illustrates the effectiveness of ingredient recognition using a huge Chinese food dataset containing photographs of complicated dish presentation.

III. METHODOLOGY

Our CNN model can predict recipes by uploading similar photographs from a 1 million-recipe dataset; however we only used 1000 recipes from this dataset because training the complete dataset with images would require a lot of memory and time to train.

This project has necessitated the development of the following modules:

Upload Recipe Dataset: All photos and recipe details can be retrieved and stored in an array using this module after it has been uploaded to a database.

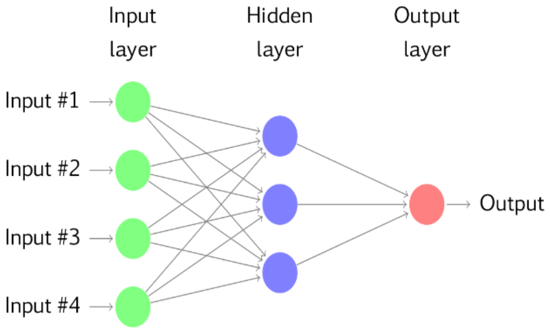
Build CNN Model: We'll use this strategy to training CNN on the recipe dataset, and then we'll feed the results back into the recipe array.

Upload Image & Predict Recipes: submit a test instance and the implementation would anticipate the recipes for that image based on that image's properties

CNN working procedure:

To show how to construct a classification model using a convolutional neural network, we will construct a 6 layer neural net that can tell one image from another. This system we're planning on building is very short and can also run on a CPU. Recurrent neural networks that really are good at classifying images have a lot too many variables and take a long time to train on a regular CPU. But our goal is to show how to use TENSORFLOW to start building a new convolutional neural network.

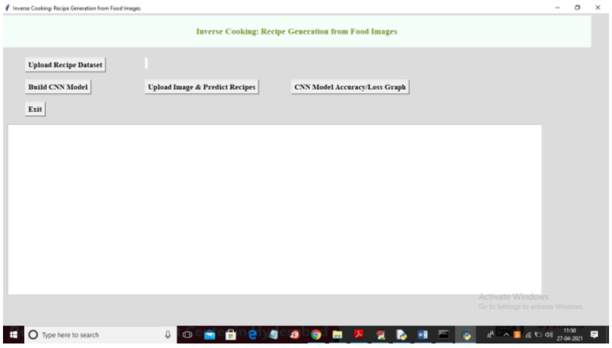
Neural Networks are basically differential equations that are used to solve a problem with optimization. [15] Neurons are the basic unit of computation in neural networks. A neuron takes in a value (let's say x), does some math on it (let's say it multiplies it by a parameter w and keeps adding another parameter b), and then comes up with a value (let's say z = wx + b). This value is sent to a non-linear function is called activation function (f) to determine a neuron's final output, or "activation." Activation functions come in many different forms. Sigmoid is a popular way to turn on a function. Sigmoid neuron is the name for a neuron whose firing function is a sigmoid function [9]. Neurons are called things like RELU and TanH based on how they are activated. There are many different kinds of neurons.A layer is what you get when users stockpile neural connections in a single line. Layers would be the next basic foundation of neural networks. See the picture with layers below.



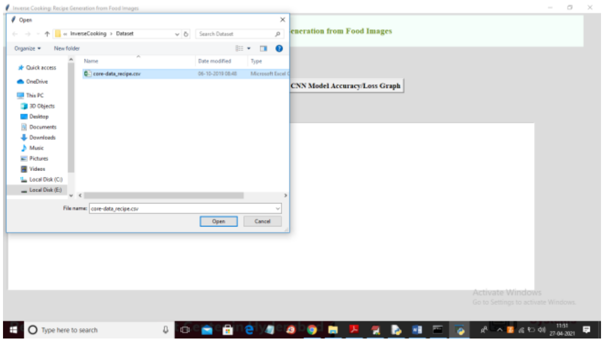
Various levels work together to find the highest rating layer, and so this procedure keeps going until there's no progress to be made.

IV. RESULT AND DISCUSSION

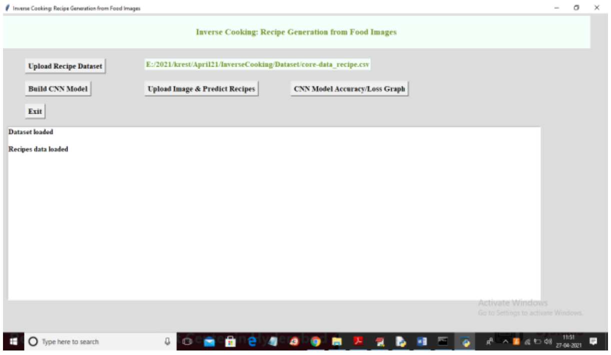
To run the project to get below result


Inverse Cooking: Recipe Generation from Food Images
Inverse Cooking: Recipe Generation from Food Images
Upload Recipe Dataset
Build CNN Model
Upload Image & Predict Recipes
CNN Model Accuracy/Loss Graph
Exit
Activate Windows
Go to Settings to activate Windows.
Type here to search
11:50
27-04-2021

The 'Upload Recipe Dataset' button can be found at the bottom of the page.

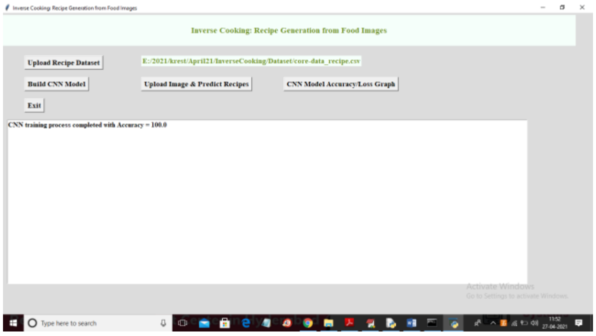


Loading a recipe dataset is done by selecting it from the drop-down menu and then clicking "open," as seen above.


Inverse Cooking: Recipe Generation from Food Images
Inverse Cooking: Recipe Generation from Food Images
Upload Recipe Dataset
E:/2021/krest/April21/InverseCooking/Dataset/core-data_recipe.csv
Build CNN Model
Upload Image & Predict Recipes
CNN Model Accuracy/Loss Graph
Exit

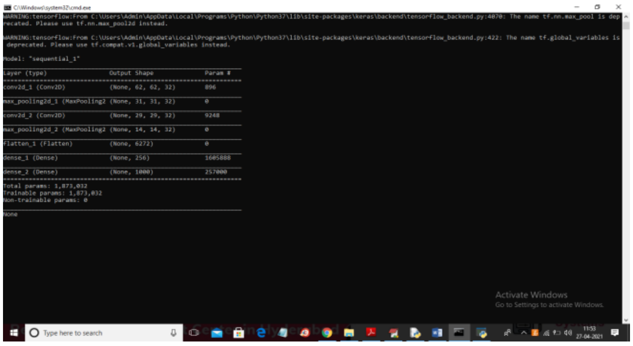
```
Dataset loaded

Recipes data loaded
```

To build a CNN using the above dataset, click on the 'Build CNN Model' button.


Inverse Cooking: Recipe Generation from Food Images
Inverse Cooking: Recipe Generation from Food Images
Upload Recipe Dataset
E:/2021/krest/April21/InverseCooking/Dataset/core-data_recipe.csv
Build CNN Model
Upload Image & Predict Recipes
CNN Model Accuracy/Loss Graph
Exit

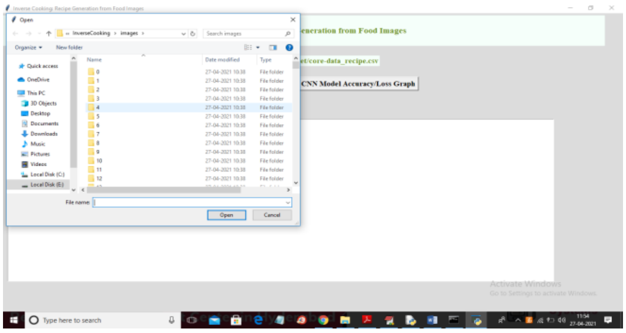
```
CNN training process completed with Accuracy = 100.0
```

Our predictions were accurate to within a single decimal point, as seen by the above output CNN model, and the CNN specifics may be seen below.

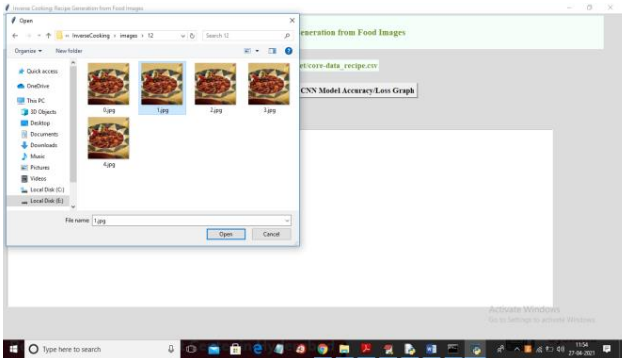


```
WARNING:tensorflow:From C:\Users\Admin\AppData\Local\Programs\Python\Python37\lib\site-packages\keras\backend\tensorflow_backend.py:4070: The name tf.nn.max_pool is dep
recated. Please use tf.nn.max_pool2d instead.

WARNING:tensorflow:From C:\Users\Admin\AppData\Local\Programs\Python\Python37\lib\site-packages\keras\backend\tensorflow_backend.py:422: The name tf.global_variables is
 deprecated. Please use tf.compat.v1.global_variables instead.

Model: "sequential_1"
_________________________________________________________________
Layer (type)                 Output Shape              Param #
=================================================================
conv2d_1 (Conv2D)            (None, 62, 62, 32)        896
_________________________________________________________________
max_pooling2d_1 (MaxPooling2 (None, 31, 31, 32)        0
_________________________________________________________________
conv2d_2 (Conv2D)            (None, 29, 29, 32)        9248
_________________________________________________________________
max_pooling2d_2 (MaxPooling2 (None, 14, 14, 32)        0
_________________________________________________________________
flatten_1 (Flatten)          (None, 6272)              0
_________________________________________________________________
dense_1 (Dense)              (None, 256)               1605888
_________________________________________________________________
dense_2 (Dense)              (None, 1000)              257000
=================================================================
Total params: 1,873,032
Trainable params: 1,873,032
Non-trainable params: 0
_________________________________________________________________
None
```

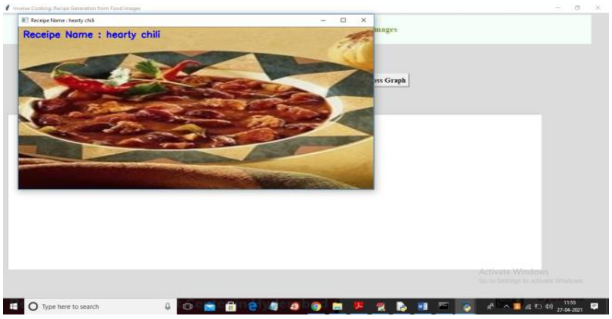
Using this dataset to train the model, we have used numerous layers and images of varied sizes, such as 62x62 and 31x31. Upload a test image by clicking the "Upload Image & Predict Recipes" button.



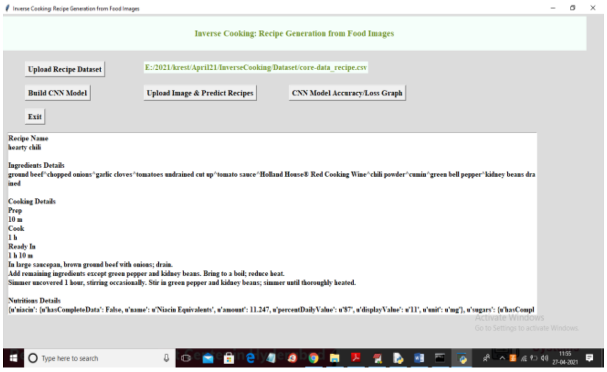
Select an image from a folder of your choice in the results shown above.



After clicking on the 'Open' button above, you will be presented with the following results.


Receipe Name : hearty chili

Close the image above to reveal the recipe details, which are labelled as 'hearty chilli' in the uploaded image.


Inverse Cooking: Recipe Generation from Food Images

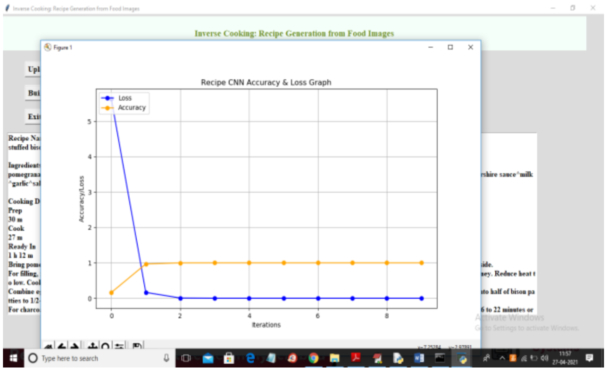
Upload Recipe Dataset
E:/2021/krest/April21/InverseCooking/Dataset/core-data_recipe.csv

Build CNN Model
Upload Image & Predict Recipes
CNN Model Accuracy/Loss Graph

Exit

```
Recipe Name
hearty chili

Ingredients Details
ground beef^chopped onions^garlic cloves^tomatoes undrained cut up^tomato sauce^Holland House® Red Cooking Wine^chili powder^cumin^green bell pepper^kidney beans dra
ined

Cooking Details
Prep
10 m
Cook
1 h
Ready In
1 h 10 m
In large saucepan, brown ground beef with onions; drain.
Add remaining ingredients except green pepper and kidney beans. Bring to a boil; reduce heat.
Simmer uncovered 1 hour, stirring occasionally. Stir in green pepper and kidney beans; simmer until thoroughly heated.

Nutritions Details
{u'niacin': {u'hasCompleteData': False, u'name': u'Niacin Equivalents', u'amount': 11.247, u'percentDailyValue': u'87', u'displayValue': u'11', u'unit': u'mg'}, u'sugars': {u'hasCompl
```

Using the 'CNN Model Accuracy/Loss Graph' button, we can view the recipe name, ingredients, cooking, and nutrition facts, as well as the ability to input any image and retrieve the recipe.



The x-axis indicates epochs, while the y-axis depicts accuracy vs. loss. The blue line depicts loss, while the orange line depicts accuracy. In the above graph, accuracy increased to 1 (100 percent) while loss decreased to 0. As the epochs increased, so did accuracy and loss. To be considered an efficient CNN model, it must have high accuracy and low loss.

V. CONCLUSION

An image-to-recipe generating method is described in this study, which starts with an image of a dish and generates an entire recipe, down to the order in which the ingredients are listed. Using food photographs as input, we were able to make predictions about ingredient combinations. It wasn't until the next step that we looked at instruction creation that we realised the need of thinking about both images and inferred constituents simultaneously [9]. User studies corroborate the challenge of this task and show that our system outperforms existing methods for converting an image into a recipe.

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. Jing-Jing Chen and Chong-Wah Ngo. Deep-based ingredient recognition for cooking recipe retrieval. In ACM Multimedia. ACM, 2016.
4. Jing-Jing Chen, Chong-Wah Ngo, and Tat-Seng Chua. Cross-modal recipe retrieval with rich food attributes. In ACM Multimedia. ACM, 2017.
5. 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.
6. Xin Chen, Hua Zhou, and Liang Diao. Chinesefoodnet: A large-scale image dataset for chinese food recognition. CoRR, abs/1705.02743, 2017.
7. Bo Dai, Dahua Lin, Raquel Urtasun, and Sanja Fidler. Towards diverse and natural image descriptions via a conditional gan. ICCV, 2017.
8. Krzysztof Dembczynski, Weiwei Cheng, and Eyke ´ Hullermeier. Bayes optimal multilabel classification via ¨ probabilistic classifier chains. In ICML, 2010.
9. Angela Fan, Mike Lewis, and Yann Dauphin. Hierarchical neural story generation. In ACL, 2018.
10. Claude Fischler. Food, self and identity. Information (International Social Science Council), 1988.
11. Jonas Gehring, Michael Auli, David Grangier, Denis Yarats, and Yann N. Dauphin. Convolutional sequence to sequence learning. CoRR, abs/1705.03122, 2017.
12. Yunchao Gong, Yangqing Jia, Thomas Leung, Alexander Toshev, and Sergey Ioffe. Deep convolutional ranking for multilabel image annotation. CoRR, abs/1312.4894, 2013.
13. Kristian J. Hammond. CHEF: A model of case-based planning. In AAAI, 1986.
14. Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Delving deep into rectifiers: Surpassing human-level performance on imagenet classification. In CVPR, 2015.
15. Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In CVPR, 2016.