Inverse Cooking: Recipe Generation From Food Images

Mellam Gayathri
PG scholar, Department of MCA, CDNR collage, Bhimavaram, Andhra Pradesh.
A.Naga Raju
(Assistant Professor), Master of Computer Applications, DNR collage, Bhimavaram, Andhra Pradesh.

Abstract:

People enjoy food photography because they appreciate food. Behind each meal there is a story described in a complex recipe and, unfortunately, by simply looking at a food image we do not have access to its preparation process. Therefore, in this paper we introduce an inverse cooking system that recreates cooking recipes given food images. Our system predicts ingredients as sets by means of a novel architecture, modeling their dependencies without imposing any order, and then generates cooking instructions by attending to both image and its inferred ingredients simultaneously. We extensively evaluate the whole system on the large-scale Recipe1M dataset and show that (1) we improve performance w.r.t. previous baselines for ingredient prediction; (2) we are able to obtain high quality recipes by leveraging both image and ingredients; (3) our system is able to produce more compelling recipes than retrieval-based approaches according to human judgment. We make code and models publicly available.

INTRODUCTION

Food is fundamental to human existence. Not only does it provide us with energy—it also defines our identity and culture [10, 34]. As the old saying goes, we are what we eat, and food related activities such as cooking, eating and talking about it take a significant portion of our daily life. Food culture has been spreading more than ever in the current digital era, with many people sharing pictures of food they are eating across social media [31]. Querying Instagram for food leads to at least 300M posts; similarly, searching for foodie results in at least 100M posts, highlighting the unquestionable value that food has in our society.

Moreover, eating patterns and cooking culture have been evolving over time. In the past, food was mostly prepared at home, but nowadays we frequently consume food prepared by third parties (e.g. takeaways, catering and restaurants). Thus, the access to detailed information about prepared food is limited and, as a consequence, it is hard to know precisely what we eat. Therefore, we argue that there is a need for inverse cooking systems, which are able to infer ingredients and cooking instructions from a prepared meal.

The last few years have witnessed outstanding improvements in visual recognition tasks such as natural image classification [47, 14], object detection [42, 41] and semantic segmentation [27, 19]. However, when comparing to natural image understanding, food recognition poses additional challenges, since food and its components have high intraclass variability and present heavy deformations that occur during the cooking process. Ingredients are frequently occluded in a cooked dish and come in a variety of colors, forms and textures. Further, visual ingredient detection requires high level reasoning and prior knowledge (e.g. cake will likely contain sugar and not salt, while croissant will presumably include butter).

Hence, food recognition challenges current computer vision systems to go beyond the merely visible, and to incorporate prior knowledge to enable high-quality structured food preparation descriptions. Previous efforts on food understanding have mainly focused on food and ingredient categorization. However, a system for comprehensive visual food recognition should not only be able to recognize the type of meal or its ingredients, but also understand its preparation process. Traditionally, the image-to-recipe problem has been formulated as a retrieval task [54, 3, 4, 45], where a recipe is retrieved from a fixed dataset based on the image similarity score in an embedding space.

The performance of such systems highly depends on the dataset size and diversity, as well as on the quality of the learned embedding. Not surprisingly, these systems fail when a matching recipe for the image query does not exist in the static dataset. An alternative to overcome the dataset constraints of retrieval systems is to formulate the image-to-recipe problem as a conditional generation one. Therefore, in this paper, we present a system that generates a cooking recipe containing a title, ingredients and cooking instructions directly from an image. Figure 1 shows an example of a generated recipe obtained with our method, which first predicts ingredients from an image and then conditions on both the image and the ingredients to generate the cooking instructions.

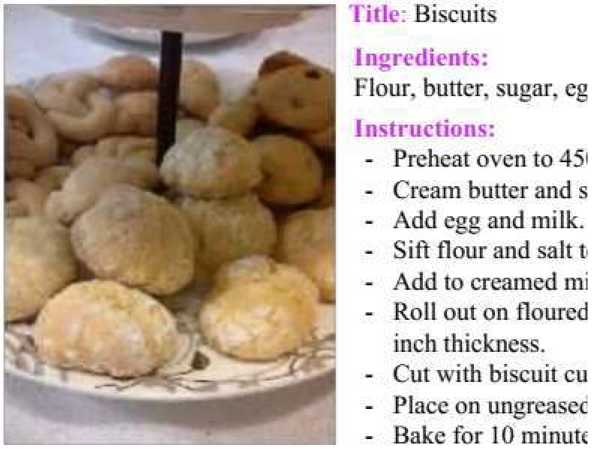
To the best of our knowledge, our system is the first to generate cooking recipes directly from food images. We pose the instruction generation problem as a sequence generation one conditioned on two modalities simultaneously, namely an image and its predicted ingredients. We formulate the ingredient prediction problem as a set prediction, exploiting their underlying structure. We model ingredient dependencies while not penalizing for prediction order, thus revising the question of whether order matters [51]. We extensively evaluate our system on the large-scale Recipe1M dataset [45] that contains images, ingredients and cooking instructions, showing satisfactory results.

More precisely, in a human evaluation study, we show that our inverse cooking system outperforms previously introduced image-to-recipe retrieval approaches by a large margin. Moreover, using a small set of images, we show that food image-to-ingredient prediction is a hard task for humans and that our approach is able to surpass them. The contributions of this paper can be summarized as:

– We present an inverse cooking system, which generates cooking instructions conditioned on an image and its ingredients, exploring different attention strategies to reason about both modalities simultaneously.

– We exhaustively study ingredients as both a list and a set, and propose a new architecture for ingredient prediction that exploits co-dependencies among ingredients without imposing order.

– By means of a user study we show that ingredient prediction is indeed a difficult task and demonstrate the superiority of our proposed system against image-to recipe retrieval approaches.



Title: Biscuits

Ingredients:

Flour, butter, sugar, eg

Instructions:

- Preheat oven to 45
- Cream butter and s
- Add egg and milk.
- Sift flour and salt t
- Add to creamed mi
- Roll out on floured inch thickness.
- Cut with biscuit cu
- Place on ungreased
- Bake for 10 minute

Figure 1: Example of a generated recipe, composed of a title, ingredients and cooking instructions.

LITEARTURE SURVEY

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

In this paper we address the problem of automatically recognizing pictured dishes. To this end, we introduce a novel method to mine discriminative parts using Random Forests (rf), which allows us to mine for parts simultaneously for all classes and to share knowledge among them. To improve efficiency of mining and classification, we only consider patches that are aligned with image superpixels, which we call components. To measure the performance of our rf component mining for food recognition, we introduce a novel and challenging dataset of 101 food categories, with 101’000 images.

With an average accuracy of 50.76%, our model outperforms alternative classification methods except for cnn, including svm classification on Improved Fisher Vectors and existing discriminative part-mining algorithms by 11.88% and 8.13%, respectively. On the challenging mit-Indoor dataset, our method compares nicely to other s-o-a component-based classification methods.

[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.

Designing powerful tools that support cooking activities has rapidly gained popularity due to the massive amounts of available data, as well as recent advances in machine learning that are capable of analysing them. In this paper, we propose a cross-modal retrieval model aligning visual and textual data (like pictures of dishes and their recipes) in a shared representation space.

We describe an effective learning scheme, capable of tackling large-scale problems, and validate it on the Recipe1M dataset containing nearly 1 million picture-recipe pairs. We show the effectiveness of our approach regarding previous state-of-the-art models and present qualitative results over computational cooking use cases. Designing powerful tools that support cooking activities has become an attractive research field in recent years due to the growing interest of users to eat home-made food and share recipes on social platforms [35].

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

Retrieving recipes corresponding to given dish pictures facilitates the estimation of nutrition facts, which is crucial to various health relevant applications. The current approaches mostly focus on recognition of food category based on global dish appearance without explicit analysis of ingredient composition. Such approaches are incapable for retrieval of recipes with unknown food categories, a problem referred to as zero-shot retrieval. On the other hand, content-based retrieval without knowledge of food categories is also difficult to attain satisfactory performance due to large visual variations in food appearance and ingredient composition.

As the number of ingredients is far less than food categories, understanding ingredients underlying dishes in principle is more scalable than recognizing every food category and thus is suitable for zero-shot retrieval. Nevertheless, ingredient recognition is a task far harder than food categorization, and this seriously challenges the feasibility of relying on them for retrieval. This paper proposes deep architectures for simultaneous learning of ingredient recognition and food categorization, by exploiting the mutual but also fuzzy relationship between them. The learnt deep features and semantic labels of ingredients are then innovatively applied for zero-shot retrieval of recipes. By experimenting on a large Chinese food dataset with images of highly complex dish appearance, this paper demonstrates the feasibility of ingredient recognition and sheds light on this zero-shot problem peculiar to cooking recipe retrieval.

[4] Jing-Jing Chen, Chong-Wah Ngo, and Tat-Seng Chua. Cross-modal recipe retrieval with rich food attributes. In ACM Multimedia. ACM, 2017.

Food is rich of visible (e.g., colour, shape) and procedural (e.g., cutting, cooking) attributes. Proper leveraging of these attributes, particularly the interplay among ingredients, cutting and cooking methods, for health-related applications has not been previously explored. This paper investigates cross-modal retrieval of recipes, specifically to retrieve a text-based recipe given a food picture as query. As similar ingredient composition can end up with wildly different dishes depending on the cooking and cutting procedures, the difficulty of retrieval originates from fine-grained recognition of rich attributes from pictures.

With a multi-task deep learning model, this paper provides insights on the feasibility of predicting ingredient, cutting and cooking attributes for food recognition and recipe retrieval. In addition, localization of ingredient regions is also possible even when region-level training examples are not provided. Experiment results validate the merit of rich attributes when comparing to the recently proposed ingredient-only retrieval techniques.

[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.

Computer-aided food identification and quantity estimation have caught more attention in recent years because of the growing concern of our health. The identification problem is usually defined as an image categorization or classification problem and several researches have been proposed. In this paper, we address the issues of feature descriptors in the food identification problem and introduce a preliminary approach for the quantity estimation using depth information. Sparse coding is utilized in the SIFT and Local binary pattern feature descriptors, and these features combined with gabor and color features are used to represent food items.

A multi-label SVM classifier is trained for each feature, and these classifiers are combined with multi-class Adaboost algorithm. For evaluation, 50 categories of worldwide food are used, and each category contains 100 photographs from different sources, such as manually taken or from Internet web albums. An overall accuracy of 68.3% is achieved, and success at top-N candidates achieved 80.6%, 84.8%, and 90.9% accuracy accordingly when N equals 2, 3, and 5, thus making mobile application practical. The experimental results show that the proposed methods greatly improve the performance of original SIFT and LBP feature descriptors. On the other hand, for quantity estimation using depth information, a straight forward method is proposed for certain food, while transparent food ingredients such as pure water and cooked rice are temporarily excluded.

of food related tasks, with special focus in image classification [26, 39, 38, 33, 6, 24, 30, 60, 16, 17].

PROPOSED METHOD

In this paper author is using RESNET101 for cuisine classification from food image and then using LSTM to predict recipe details. All food contains its own ingredients and recipe details so just by seeing food images it’s not possible to identify Cuisine name and its ingredients. So author of this paper using RESNET and LSTM for cuisine classification and to predict ingredients and recipe details. RESNET

To train both algorithms author using 1 Million recipe dataset from KAGGLE website. Author has compare performance of RESNET101 with SVM and XGBOOST, Cat Boost and Light GBM. Cat Boost and Light GBM taking more execution time so we have implemented Resnet101, SVM and XGBOOST.

To implement this project we have designed following modules

1) Upload Recipe Dataset: using this module we will upload dataset and then read all images and its recipe details and then build TF-IDF vector which contains average frequency of each word occurrence and LSTM will get trained on this vector and then retrieve all vector which is matching with predicted cuisine food image
2) Build Resnet101 Model: using this module we will train Resnet101 on images and LSTM on TF-IDF vector and then calculate its prediction accuracy
3) Train SVM & XGBoost Algorithms: using this module we will train SVM and XGBOOST algorithm and then calculate its prediction accuracy
4) Upload Image & Predict Recipes: using this module we will upload test food image and then Resnet101 will classify cuisine and LSTM will predict ingredients and recipe details.
5) Comparison Graph: using this module we will plot accuracy graph between all algorithms

RESULT

To run project double click on ‘run.bat’ file to get below screen

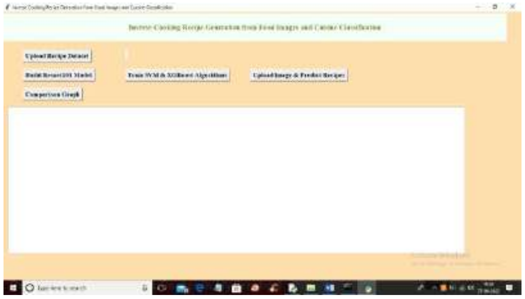

Inverse Cooking Recipe Generation from Food Images and Cuisine Classification
Upload Recipe Dataset
Build Resnet101 Model
Train SVM & XGBoost Algorithms
Upload Image & Predict Recipe
Comparison Graph

Fig. 5 Run.bat file

In above screen click on ‘Upload Recipe Dataset’ button to upload recipe dataset and get below output

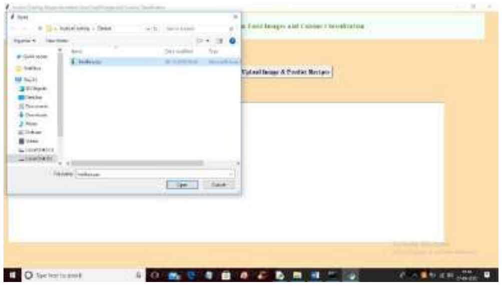
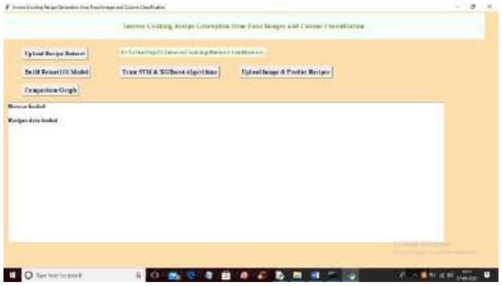


Fig.6 Upload the dataset

In above screen selecting and uploading recipe dataset file and then click on ‘Open’ button to load dataset and get below output

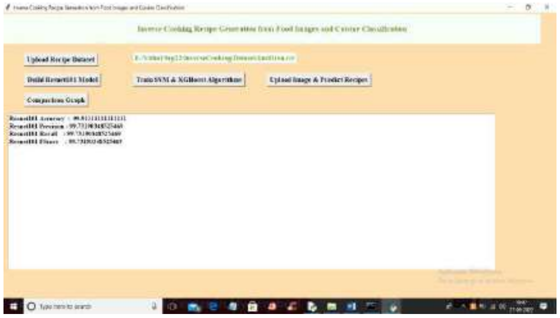

Inverse Cooking Recipe Generation from Food Images and Cuisine Classification
Upload Recipe Dataset
Build Resnet101 Model
Train SVM & XGBoost Algorithms
Upload Image & Predict Recipe
Comparison Graph

```
Dataset loaded

Recipes data loaded
```

Fig.7 Select and upload the recipe dataset

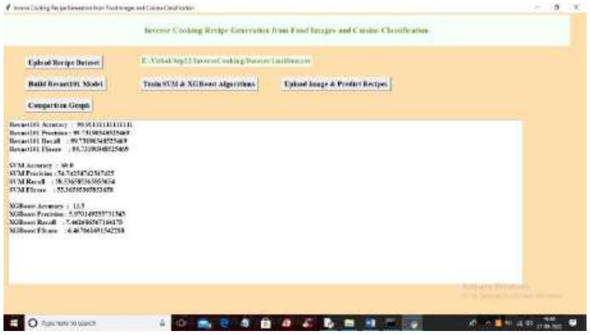
In above screen dataset loaded and now click on ‘Build Resnet101 Model’ button to train Resnet101 and get below output


Inverse Cooking Recipe Generation from Food Images and Cuisine Classification
Upload Recipe Dataset
Build Resnet101 Model
Train SVM & XGBoost Algorithms
Upload Image & Predict Recipe
Comparison Graph

```
Resnet101 Accuracy  : 99.91111111111111
Resnet101 Precision : 99.73190348525469
Resnet101 Recall    : 99.73190348525469
Resnet101 FScore    : 99.73190348525469
```

Fig.8 Build Resnet101 Model

In above screen with Resnet101 training completed and we got 99.91% accuracy and now click on ‘Train SVM & XGBoost Algorithms’ button to train both algorithms and get below output


Inverse Cooking Recipe Generation from Food Images and Cuisine Classification
Upload Recipe Dataset
Build Resnet101 Model
Train SVM & XGBoost Algorithms
Upload Image & Predict Recipe
Comparison Graph

```
Resnet101 Accuracy  : 99.91111111111111
Resnet101 Precision : 99.73190348525469
Resnet101 Recall    : 99.73190348525469
Resnet101 FScore    : 99.73190348525469

SVM Accuracy  : 60.0
SVM Precision : 54.74254742547425
SVM Recall    : 58.53658536585365
SVM FScore    : 55.36585365853658

XGBoost Accuracy  : 12.5
XGBoost Precision : 5.970149253731343
XGBoost Recall    : 5.462686567164179
XGBoost FScore    : 6.467661691542288
```

Fig.10 Train SVM & XGBoost Algorithms’

In above screen SVM and XGBOOST training completed and with SVM we got 60% accuracy and with XGBOOST we got 12.5% accuracy and in above screen we can see in all algorithms Resnet101 accuracy is high. Now click on ‘Upload Image & Predict Recipe’ button to upload image and get below output

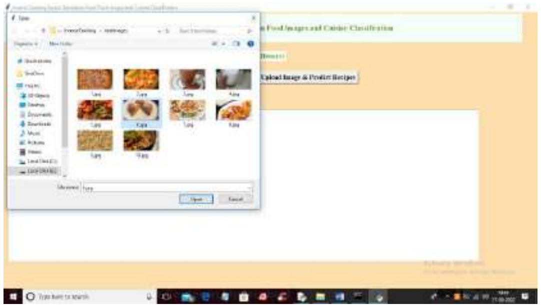
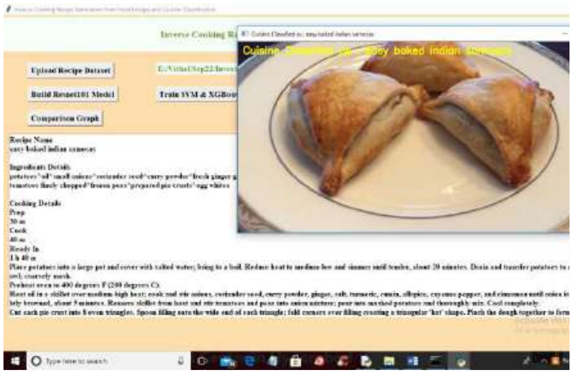


Fig.11 selecting and uploading ‘6.jpg’ file

In above screen selecting and uploading ‘6.jpg’ file and then click on ‘Open’ button to get below output



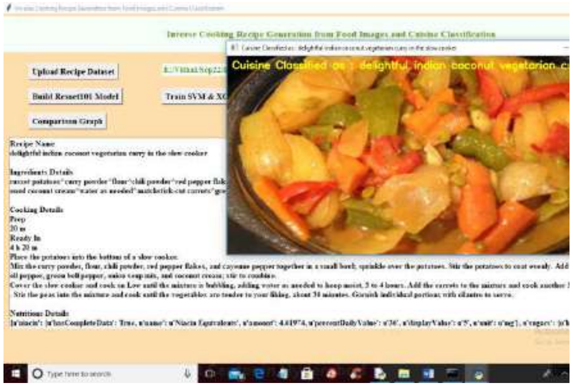
Recipe Name
easy baked indian samosas

Ingredients Details
potatoes''oil''small onions''coriander seed''curry powder''fresh ginger
tomatoes finely chopped''frozen peas''prepared pie crusts''egg whites

Cooking Details
Prep
30 m
Cook
40 m
Ready In
1 h 40 m
Place potatoes into a large pot and cover with salted water; bring to a boil. Reduce heat to medium-low and simmer until tender, about 20 minutes. Drain and transfer potatoes to a
owl; coarsely mash.
Preheat oven to 400 degrees F (200 degrees C).
Heat oil in a skillet over medium-high heat; cook and stir onions, coriander seed, curry powder, ginger, salt, turmeric, cumin, allspice, cayenne pepper, and cinnamon until onion is
htly browned, about 5 minutes. Remove skillet from heat and stir tomatoes and peas into onion mixture; pour into mashed potatoes and thoroughly mix. Cool completely.
Cut each pie crust into 8 even triangles. Spoon filling onto the wide end of each triangle; fold corners over filling creating a triangular 'hat' shape. Pinch the dough together to form

Fig.12 cuisine classification in yellow colour text

In above screen on image we can see cuisine classification in yellow colour text and in text area also we can see recipe name with ingredients and making (recipe) details. Now upload another images


Inverse Cooking Recipe Generation from Food Images and Cuisine Classification
Cuisine Classified as : delightful indian coconut vegetarian curry in the slow cooker
Cuisine Classified as : delightful indian coconut vegetarian c
Upload Recipe Dataset
Build Resnet101 Model
Train SVM & XG
Comparison Graph
Type here to search

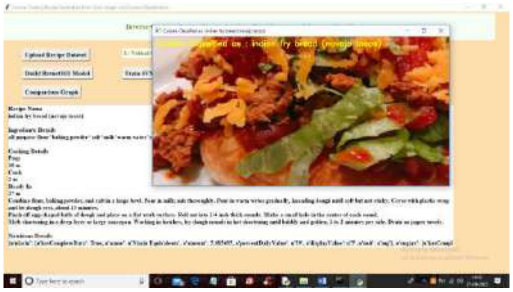
Recipe Name
delightful indian coconut vegetarian curry in the slow cooker

Ingredients Details
russet potatoes''curry powder''flour''chili powder''red pepper flak
sweet coconut cream''water as needed''matchstick-cut carrots''gr

Cooking Details
Prep
20 m
Ready In
4 h 20 m
Place the potatoes into the bottom of a slow cooker.
Mix the curry powder, flour, chili powder, red pepper flakes, and cayenne pepper together in a small bowl; sprinkle over the potatoes. Stir the potatoes to coat evenly. Add
ell pepper, green bell pepper, onion soup mix, and coconut cream; stir to combine.
Cover the slow cooker and cook on Low until the mixture is bubbling, adding water as needed to keep moist, 3 to 4 hours. Add the carrots to the mixture and cook another
. Stir the peas into the mixture and cook until the vegetables are tender to your liking, about 30 minutes. Garnish individual portions with cilantro to serve.

Nutrition Details
{u'niacin': {u'hasCompleteData': True, u'name': u'Niacin Equivalents', u'amount': 4.41974, u'percentDailyValue': u'34', u'displayValue': u'5', u'unit': u'mg'}, u'sugars': {u'h

Fig.13 Recipe details



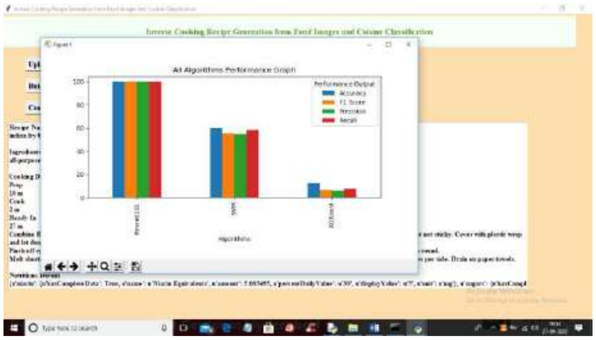
Recipe Name
indian fry bread (navajo tacos)

Ingredients Details
all-purpose flour''baking powder''salt''milk''warm water

Cooking Details
Prep
10 m
Cook
2 m
Ready In
27 m

Fig.14 Recipe details

Similarly you can upload other images and get recipe and cuisine classification details. Now click on 'Comparison Graph' button to get below graph



Nutrition Details
{u'niacin': {u'hasCompleteData': True, u'name': u'Niacin Equivalents', u'amount': 5.693455, u'percentDailyValue': u'39', u'displayValue': u'7', u'unit': u'mg'}, u'sugars': {u'hasCompl

Fig.15 Comparison Graph

In above graph x-axis represents algorithm names with each metric like accuracy, precision in different colour bar and y-axis represents values and in above graph we can see compare to all algorithm RESNET101 prediction or classification accuracy is high

CONCLUSION

In this paper, we introduced an image-to-recipe generation system, which takes a food image and produces a recipe consisting of a title, ingredients and sequence of cooking instructions. We first predicted sets of ingredients from food images, showing that modelling dependencies matters. Then, we explored instruction generation conditioned on images and inferred ingredients, highlighting the importance of reasoning about both modalities at the same time. Finally, user study results confirm the difficulty of the task, and demonstrate the superiority of our system against state of-the-art image-to-recipe retrieval approaches.

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.

[16] Luis Herranz, Shuqiang Jiang, and Ruihan Xu. Modeling restaurant context for food recognition. IEEE Transactions on Multimedia, 2017.

[17] Shota Horiguchi, Sosuke Amano, Makoto Ogawa, and Kiyoharu Aizawa. Personalized classifier for food image recognition. IEEE Transactions on Multimedia, 2018.

[18] 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.