IJARST
International Journal For Advanced Research In Science & Technology
A peer reviewed international journal
www.ijarst.in
ISSN: 2457-0362

INVERSE COOKING RECIPE GENERATION FROM FOOD IMAGES

Under the Guidance of
Mr.V.Rama Krishna
Asst.Professor
Department Of Computer Science and Engineering
Anurag Group Of Institutions Hyderabad,India

Parepalli Tagore
Department Of Computer Science
And Engineering
Anurag Group Of Institutions
Hyderabad,India
parepallitagore@gmail.com

Sasanapuri Sirila
Department Of Computer Science
And Engineering
Anurag Group Of Institutions
Hyderabad,India
sirila.sasanapuri@gmail.com

Veeramalla Satwik
Department Of Computer Science
And Engineering
Anurag Group Of Institutions
Hyderabad,India
satwikgoud002@gmail.com

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.. 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.

Keywords- *Inverse cooking, Recipe1M dataset.*

1. 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]. In the past, food was mostly prepared at home, but nowadays we frequently consume food prepared by thirdparties (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.

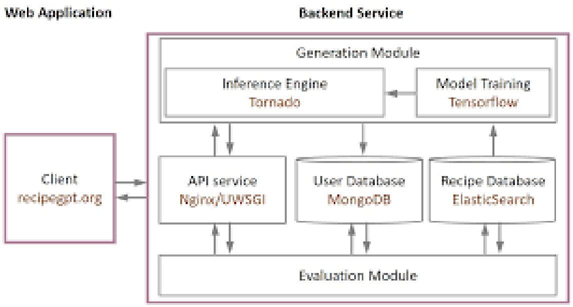


Fig.1: Example figure

The last few years have witnessed outstanding improvements in visual recognition tasks such as natural image classification, object detection and semantic segmentation. 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, 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.

2. LITERATURE REVIEW

2.1 Food-101–mining discriminative components with random forests [1] :

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.

2.2 Deep-based ingredient recognition for cooking recipe retrieval [2]:

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.

2.3 Automatic chinese food identification and quantity estimation [3] :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.

2.4 Chinesefoodnet: A large-scale image dataset for chinese food recognitio [4]:

In this paper, we introduce a new and challenging large-scale food image dataset called “ChineseFoodNet”, which aims to automatically recognizing pictured Chinese dishes. Most of the existing food image datasets collected food images either from recipe pictures or selfie. In our dataset, images of each food category of our dataset consists of not only web recipe and menu pictures but photos taken from real dishes, recipe and menu as well. ChineseFoodNet contains over 180,000 food photos of 208 categories, with each category covering a large variations in presentations of same Chinese food. We present our efforts to build this large-scale image dataset, including food category selection, data collection, and data clean and label, in particular how to use machine learning methods to reduce manual labeling work that is an expensive process.We share a detailed benchmark of several state-of-the-art deep convolutional neural networks (CNNs) on ChineseFoodNet. We further propose a novel two-step data fusion approach referred as “TastyNet”, which combines prediction results from different CNNs with voting method. Our proposed approach achieves top-1 accuracies of 81.43% on the validation set and 81.55% on the test set, respectively.

3. IMPLEMENTATION

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 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 data.

Disadvantages of existing system:

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.

In this project we are training CNN with recipe details and images and this model can be used to predict recipe by uploading related images and we used 1 million recipe dataset and from this dataset we used 1000 recipes as training entire dataset with images will take lots of memory and hours of time train CNN model.

Advantages of proposed system:

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.

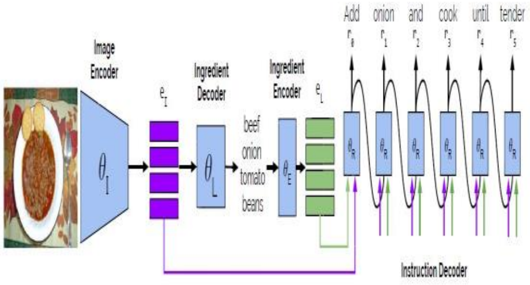


Fig.1: System architecture

To implement this project we have designed following modules

1) Upload Recipe Dataset: Using this module we will upload dataset to application and then read all images and recipes details and then store them in array

2) Build CNN Model: Using this model we will entire recipe array and then input those details to CNN model to train CNN on recipe dataset

3) Upload Image & Predict Recipes: using this module we will upload test image and the application will predict recipe for that image.

4. ALGORITHM

CNN:

Artificial Intelligence has been witnessing a monumental growth in bridging the gap between the capabilities of humans and machines. Researchers and enthusiasts alike, work on numerous aspects of the field to make amazing things happen. One of many such areas is the domain of Computer Vision. The agenda for this field is to enable machines to view the world as humans do, perceive it in a similar manner and even use the knowledge for a multitude of tasks such as Image & Video recognition, Image Analysis & Classification, Media Recreation, Recommendation Systems, Natural Language Processing, etc. The advancements in Computer Vision with Deep Learning has been constructed and perfected with time, primarily over one particular algorithm — a Convolutional Neural Network.

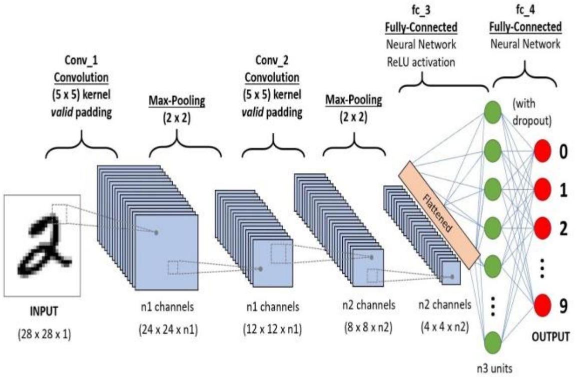


Fig.3: CNN model

A Convolutional Neural Network (ConvNet/CNN) is a Deep Learning algorithm which 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. The pre-processing required in a ConvNet is much lower as compared to other classification algorithms. While in primitive methods filters are hand-engineered, with enough training, ConvNets have the ability to learn these filters/characteristics. The architecture of a ConvNet is analogous to that of the connectivity pattern of Neurons in the Human Brain and was inspired by the organization of the Visual Cortex. Individual neurons respond to stimuli only in a restricted region of the visual field known as the Receptive Field. A collection of such fields overlap to cover the entire visual area.

Adding a Fully-Connected layer is a (usually) cheap way of learning non-linear combinations of the high-level features as represented by the output of the convolutional layer. The Fully-Connected layer is learning a possibly non-linear function in that space. Now that we have converted our input image into a suitable form for our Multi-Level Perceptron, we shall flatten the image into a column vector. The flattened output is fed to a feed-forward neural network and backpropagation applied to every iteration of training. Over a series of epochs, the model is able to distinguish between dominating and certain low-level features in images and classify them using the Softmax Classification technique.

5. EXPERIMENTAL RESULTS

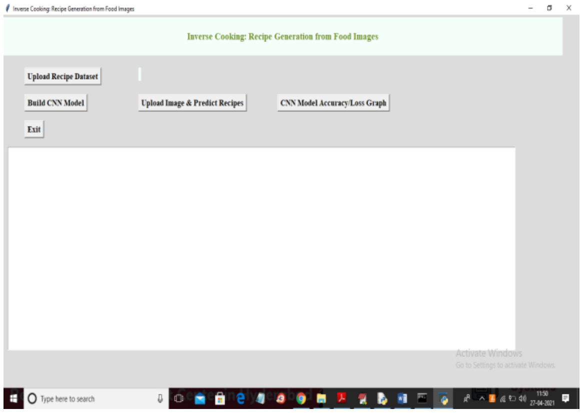

Inverse Cooking: Recipe Generation from Food Images
Inverse Cooking: Recipe Generation from Food Images
Upload Recipe Dataset
Build CNN Model
Upload Image & Predict Recipes
CNN Model Accuracy/Loss Graph
Exit
Activate Windows
Go to Settings to activate Windows.
Type here to search
11:50
27-04-2021

Fig.4: Home screen

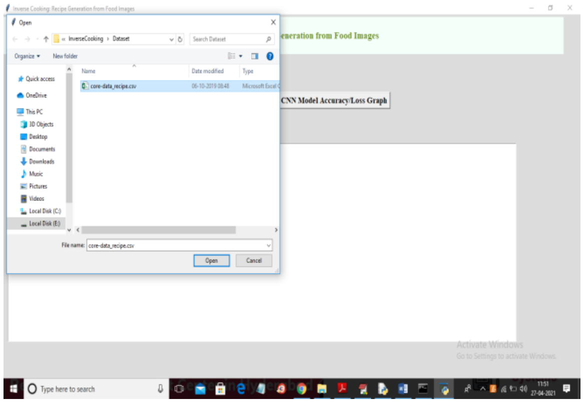
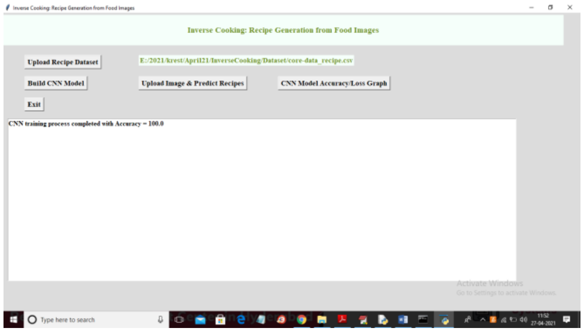


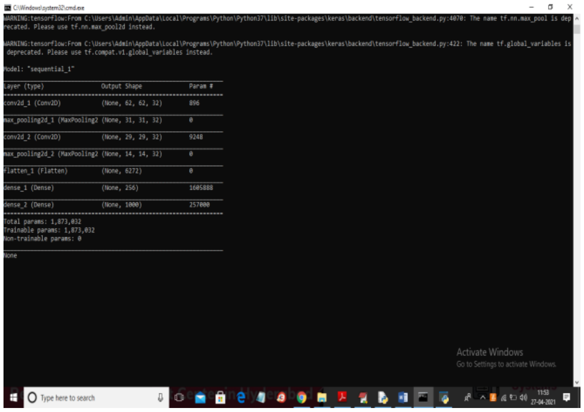
Fig.5: Upload recipe dataset


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
11:52
27-04-2021

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

CNN training process completed with Accuracy = 100.0

Fig.6: Build CNN model


C:\Windows\system32\cmd.exe
Activate Windows
Go to Settings to activate Windows.
Type here to search
11:53
27-04-2021

```
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
```

Fig.7: Train model

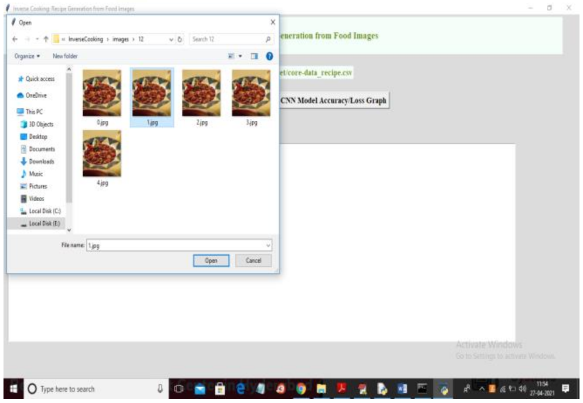


Fig.8: Upload image & predict recipes

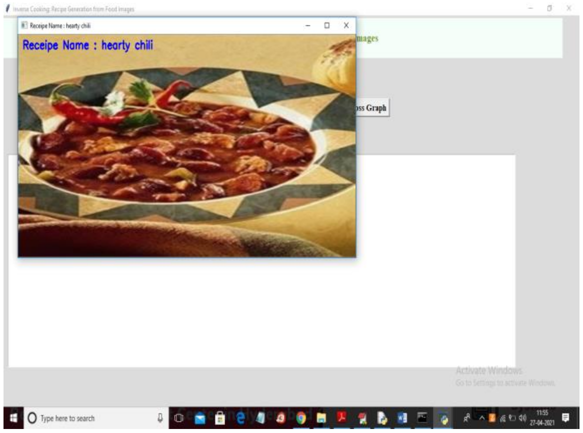
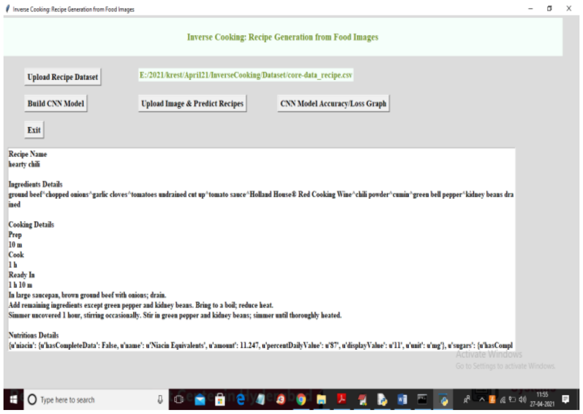

Inverse Cooking: Recipe Generation from Food Images
Receipe Name : hearty chili
Receipe Name : hearty chili
Type here to search

Fig.9: prediction 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
Type here to search

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

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 drained

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

Fig.10: Information about result

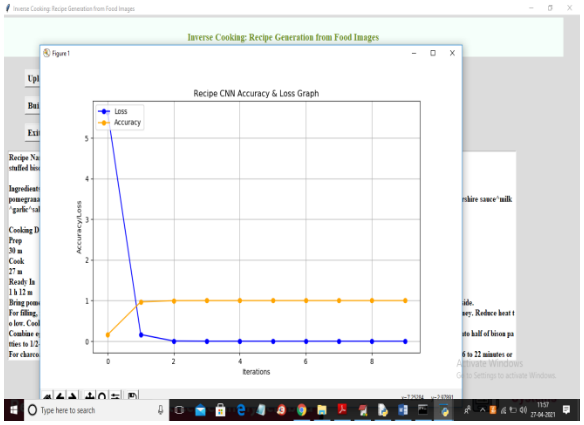


Fig.11: CNN accuracy loss graph

6. 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 modeling 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, R´emi 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 Dembczy´nski, Weiwei Cheng, and Eyke H¨ullermeier. 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.