Food Recognition for Dietary Assessment Using Deep Convolutional Neural Networks

Stergios Christodoulidis[1, 2], Marios Anthimopoulos[1,3], Stavroula Mougiakakou[1,4]

[1] ARTORG Center for Biomedical Engineering Research, University of Bern, Switzerland
[2] Graduate school of Cellular and Biomedical Sciences, University of Bern, Switzerland
[3] Department of Emergency Medicine, Bern University Hospital, Switzerland
[4] Department of Endocrinology, Diabetes and Clinical Nutrition, Bern University Hospital, Switzerland

DIABETES TECHNOLOGY RESEARCH GROUP
UNIVERSITÄT BERN
ARTORG CENTER BIOMEDICAL ENGINEERING RESEARCH

Introduction

Diet management is a key factor for the prevention and treatment of diet-related chronic diseases. Computer vision systems aim to provide automated food intake assessment using meal images. We propose a method for the recognition of food items in meal images using a deep convolutional neural network (CNN) followed by a voting scheme. Our approach exploits the outstanding descriptive ability of a CNN, while the patch-wise model allows the generation of sufficient training samples, provides additional spatial flexibility for the recognition and ignores background pixels.

Materials & Methods

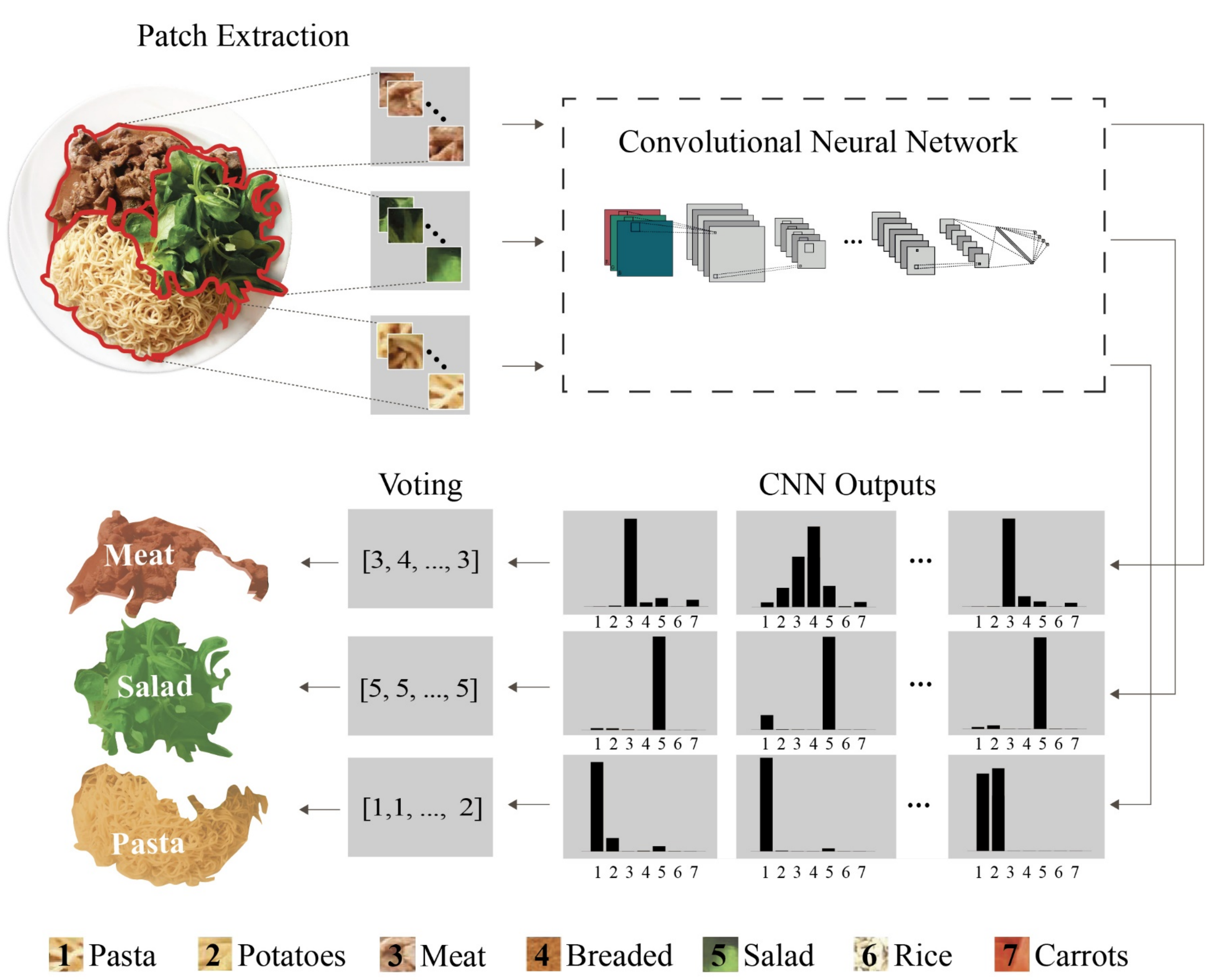

Figure 1 – The proposed system overview

The proposed system recognizes already segmented food items using an ensemble learning model (**Figure 1**).

Materials:
- 12,500 patches were extracted from 573 food items
- Seven (7) broad food classes: pasta, potatoes, meat, breaded food, rice, green salad and carrots.

Preprocessing:
- Zero mean normalization for every patch
- Data augmentation with 16 label preserving transformations leading to 200,000 patches

Network Training: A deep CNN was trained with:
- Four convolutional layers with 5×5 kernels
- Two fully connected layers at the end
- Rectified linear units (ReLU) as activation functions
- Dropout in the fully connected layers

Food Recognition: For each food item:
- Image patches are extracted on a grid
- Patches are preprocessed and fed to the CNN
- The most frequent class is assigned to the food item

Results

Evaluation:
- 60% of the data was used for training 20% for validation and 20% for testing.
- The best configuration of the network was identified by a trial and error procedure
- The performance was assessed in terms of the average F-score over the different classes on a patch basis (pF_{avg}) and food item basis (iF_{avg})
- Average classification times for food items are also given

The overall iF_{avg} of the proposed system was **84.9%**. **Table 1** provides the accuracy of the patch-wise classification for the different CNN architectures. **Table 2** shows the results of the food-item classification for a number of different schemes as well as a comparison with a method from the literature. **Figure 2** illustrates the learned kernels of the first layer convolutional layer of the CNN.

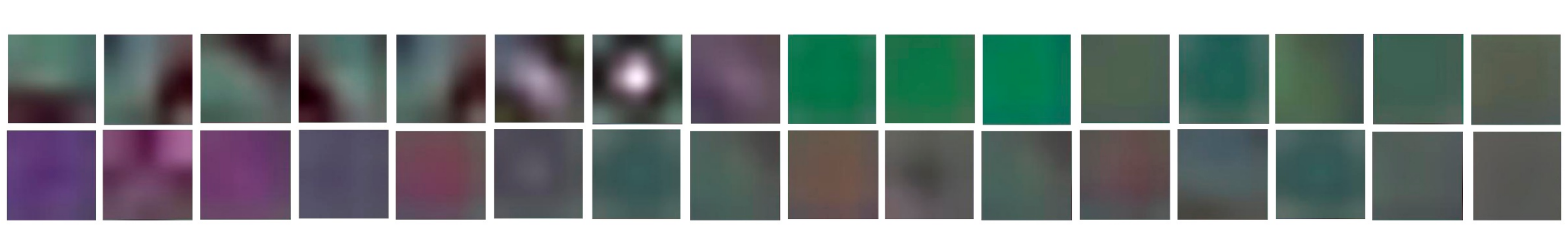

Figure 2 – The kernels from the first layer of the proposed CNN

Table 1 – Results for the different architectures investigated

CNN architecture	pF_{avg}
32cp – 32cp – 128fc – 7fc	66.5
32cp – 32cp – 64cp – 128fc – 7fc	68.7
32cp – 32cp – 32cp – 64cp – 128fc – 7fc	69.5
32cp – 32cp – 32cp – 64cp – 64cp – 128fc – 7fc	67.1
32cp – 32cp – 32cp – 64cp – 128fc – 7fc + LRN	70.4
32cp – 32cp – 32cp – 64cp – 128fc – 7fc + Dropout	**71.79**
32cp – 32cp – 32cp – 64cp – 128fc – 7fc + LRN +Dropout	71.28

Table 2 – Results of the proposed method for different voting schemes and variants compared to a method from the literature

Classification Method	Accuracy	iF_{avg}	Time (sec/item)
CNN+Weighted voting+step=16	84.6	82.8	0.28
CNN+Max voting+step=32	83.5	81.4	0.11
CNN+Max voting+step=16	**84.9**	**82.7**	**0.28**
CNN+Max voting+step=8	84.7	82.5	0.92
Learned histogram+LBP+SVM[1]	82.2	79.7	0.1

[1] M. Anthimopoulos, J. Dehais, P. Diem, S. Mougiakakou, "Segmentation and recognition of multi-food meal images for carbohydrate counting", 2013 IEEE 13th International Conference on Bioinformatics and Bioengineering (BIBE), 10-13 Nov. 2013

Conclusions

We proposed a method for the recognition of already segmented food items using a CNN. The classification is applied in a patch-wise manner and a voting technique was used to determine the class of each food item. Future work should include a more thorough investigation on the optimal architecture as well as the training parameters of the network.

Acknowledgements

The work was funded in part by the Bern University Hospital "Inselspital" and the European Union Seventh Framework Programme (FP7-PEOPLE-2011-IAPP) under grant agreement n° 286408 [www.gocarb.eu]. The consortium would like to thank Swisscom for generously providing the mobile devices and services.

INSELSPITAL
UNIVERSITÄTSSPITAL BERN
HOPITAL UNIVERSITAIRE DE BERNE
BERN UNIVERSITY HOSPITAL
swisscom