Food Ingredients Recognition through Multi-label Learning

Marc Bolaños, Aina Ferrà and Petia Radeva

marc.bolanos@ub.edu aferrama10.alumnes@ub.edu petia.ivanova@ub.edu

UNIVERSITAT DE BARCELONA

CVC Centre de Visió per Computador

MADiMa 2017

1 Objective

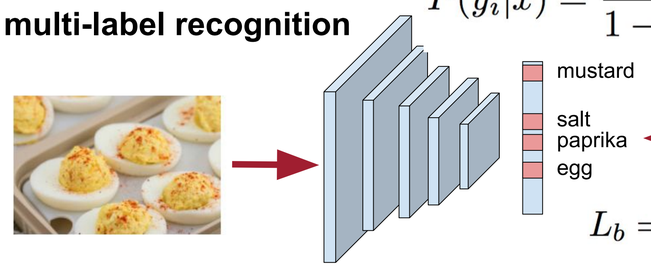
- We tackle the problem of food ingredients recognition as a multi-label learning [1] problem.
- Propose a method for adapting a highly performing state of the art CNN and act as a multi-label predictor for learning recipes in terms of their list of ingredients.
- Finally, we visualize the activations and prove that the neurons are specialized on finding certain ingredients.

2 Deep Multi-Ingredients Recognition

Our proposal: Adaptation for multi-label recognition

Sigmoid

$$P(y_i|x) = \frac{1}{1-\exp^{-f(x)_i}}$$



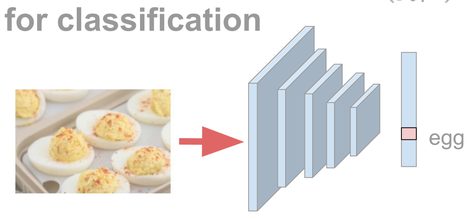
Binary cross-entropy [2]

$$L_b = -\sum_x \sum_i^N (\hat{y}_x^i \cdot log(P(y_i|x)) + (1-\hat{y}_x^i) \cdot log(1-P(y_i|x)))$$

Conventional approach: CNN for classification

Softmax

$$P(y_i|x) = \frac{\exp^{f(x)_i}}{\sum_i \exp^{f(x)_j}}$$



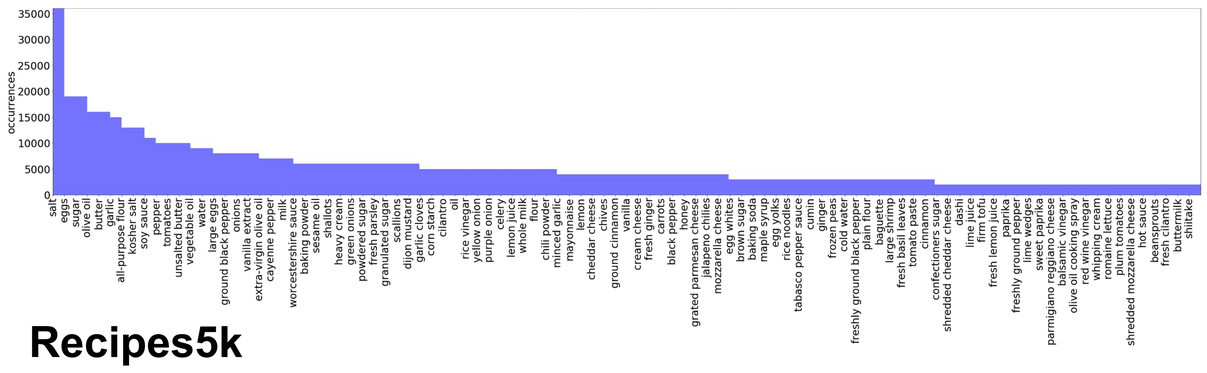
Categorical cross-entropy

$$L_c = -\sum_x \log(P(\hat{y}_x|x))$$

3 Datasets

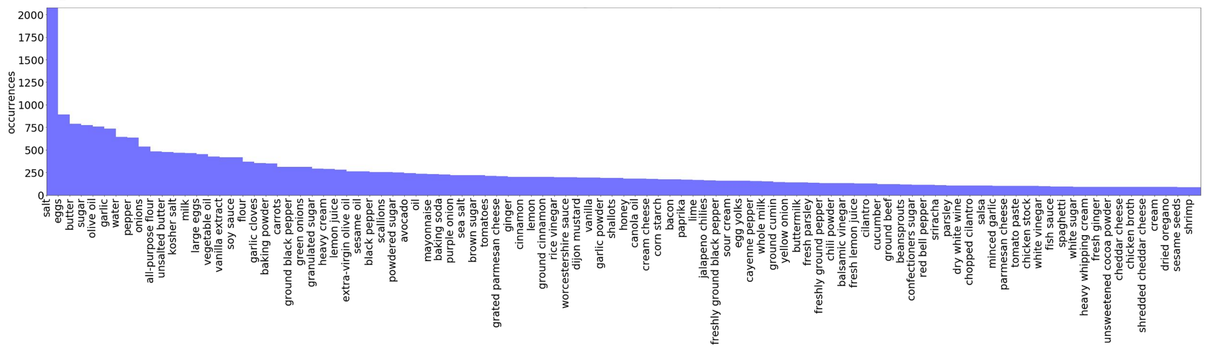
Ingredients101

Complementary to Food101 [3]. A reference recipe was downloaded from Yummly.com for each class, summing up 446 ingredients.



Recipes5k

Around 50 different recipes were downloaded for each class in Food101 from Yummly.com. Composed of nearly 5,000 images and 3,213 unique ingredients.



Ingredients Simplification

Many of the 3,213 raw ingredients are sub-classes (e.g. 'sliced tomato' or 'tomato sauce') of more general versions (e.g. 'tomato').

We propose a simplified version by applying a simple removal of overly-descriptive particles (e.g. 'sliced' or 'sauce'), resulting in 1,013 ingredients used for additional evaluation.

4 Results

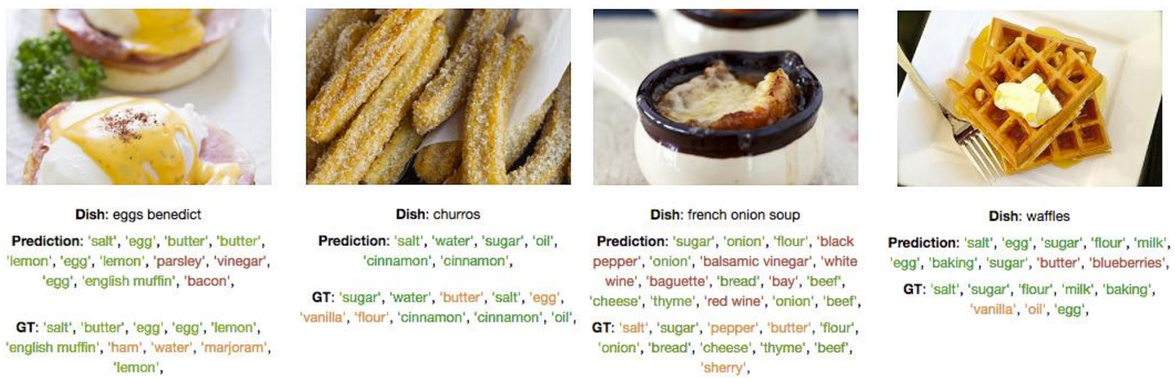
Ingredients101

	Validation			Test		
	Prec	Rec	F_1	Prec	Rec	F_1
Random prediction	2.05	2.01	2.03	2.06	2.01	2.04
InceptionV3 + Ingredients101	80.86	72.12	76.24	83.51	**76.87**	80.06
ResNet50 + Ingredients101	84.80	67.62	75.24	**88.11**	73.45	**80.11**

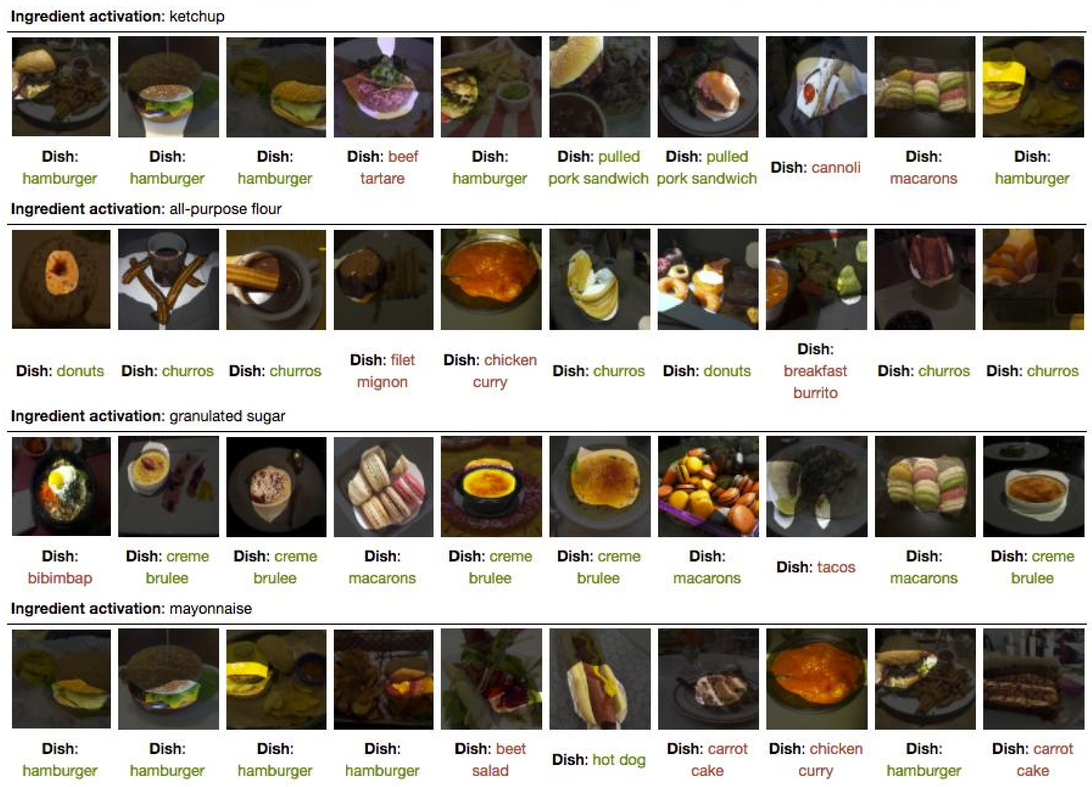
Recipes5k

	Validation			Test		
	Prec	Rec	F_1	Prec	Rec	F_1
Random prediction	0.33	0.32	0.33	0.54	0.53	0.53
InceptionV3 + Ingredients101				23.80	18.24	20.66
ResNet50 + Ingredients101				26.28	16.85	20.54
InceptionV3 + Recipes5k	36.18	20.69	26.32	35.47	**21.00**	**26.38**
ResNet50 + Recipes5k	38.41	19.67	26.02	**38.93**	19.57	26.05
Random prediction	6.27	6.29	6.28	6.14	6.24	6.19
InceptionV3 + Ingredients101				44.01	34.04	38.39
ResNet50 + Ingredients101				47.53	30.91	37.46
InceptionV3 + Recipes5k	56.77	31.40	40.44	55.37	31.52	40.18
ResNet50 + Recipes5k	56.73	28.07	37.56	**58.55**	28.49	38.33
InceptionV3 + Recipes5k simplified	53.91	42.13	47.30	53.43	**42.77**	**47.51**

SIMPLIFIED INGREDIENTS



5 Neurons' Activations



6 Conclusions

The proposed model and the two datasets published offer very promising results for the multi-label problem of food ingredients recognition. Our proposal allows to obtain great generalization results on unseen recipes and sets the basis for applying further, more detailed food analysis methods

7 Bibliography

[1] Grigorios Tsoumakas and Ioannis Katakis. Multi-label classification: An overview. International Journal of Data Warehousing and Mining, 3(3), 2006.

[2] Andreas Buja, Werner Stuetzle, and Yi Shen. Loss functions for binary class probability estimation and classification: Structure and applications. Working draft, November, 2005.

[3] Lukas Bossard, Matthieu Guillaumin, and Luc Van Gool. Food-101–mining discriminative components with random forests. In European Conference on Computer Vision, pages 446–461. Springer, 2014.

Acknowledgments: This work was partially founded by TIN2015-66951-C2, SGR 1219, ICREA Academia 2014, Grant FPU15/01347.
We acknowledge NVIDIA for the donation of a GPU.