Uncertainty modeling within an End-to-end framework for Food Image Analysis

Petia Radeva,

Collaboration with:
Eduardo Aguilar, Marc Bolaños,
Bhalaji Nagarajan, Rupali Khatun

University of Barcelona &
Computer Vision Center

radevap@gmail.com

Contents

Why food recognition?

"Camera eats first"

180M #food
90/minute

54% take picture
39% post it

What is the gender?

“AI: More than Human” Forum, Groningen

Why is the food recognition a challenge?

Motivation

Food Analysis Problems

Ingredients

- Intra-class variability
- Inter-class similarity

Intra-class variability example: Apple. Image source: Recipes5k

Inter-class similarity example: Tomato sauce and Curry sauce. Image source: Recipes5k

Decreasement in Precision

Are we able to recognize thousands of dishes?

- 79% on UECFOOD
- 44% on ChinaFood1000
- How to achieve scalability?

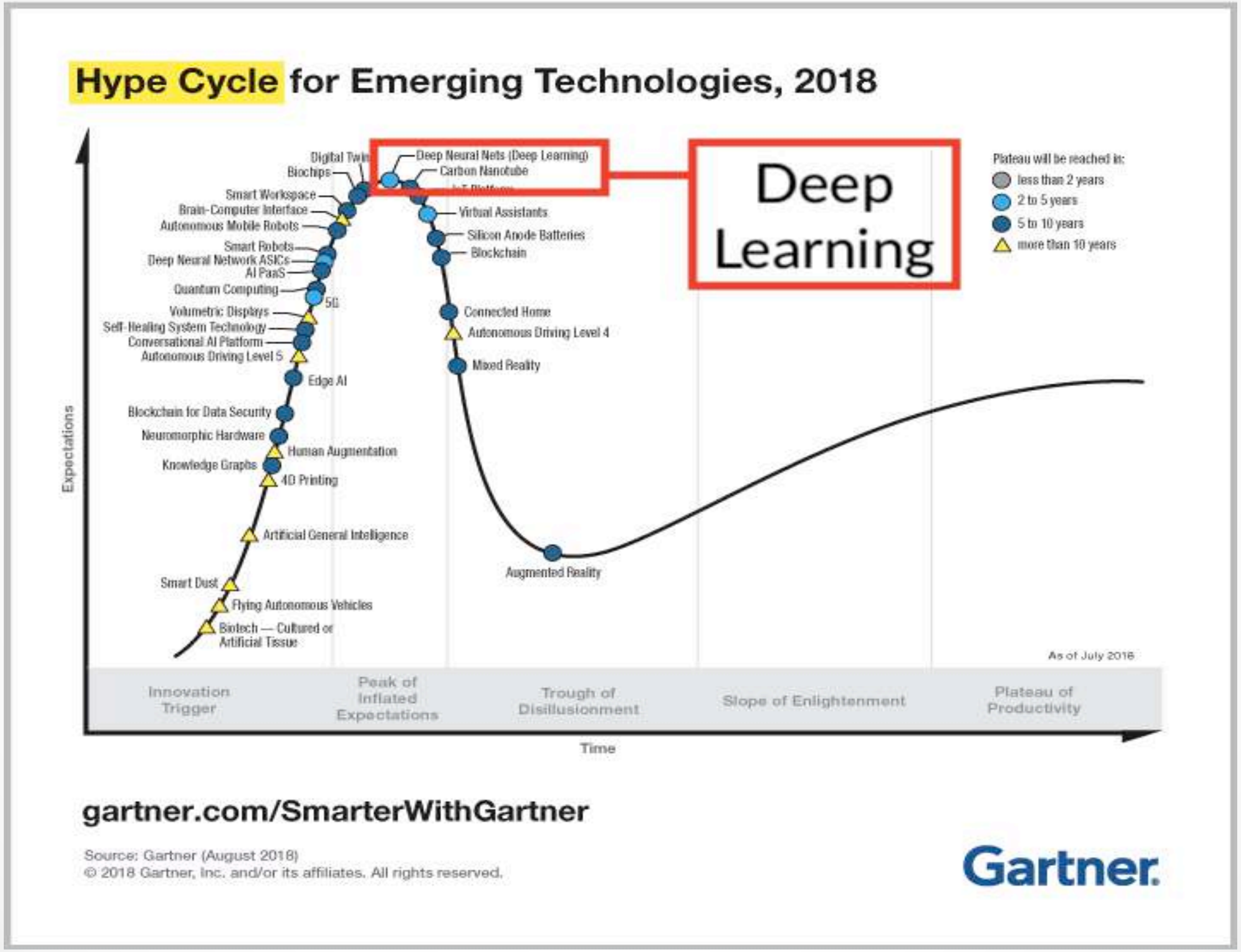
Contents

- The food image problem
- Deep learning for food analysis
- Uncertainty modelling
- Multi-task food learning with aleatoric uncertainty
- Hierarchical food recognition with epistemic uncertainty
- Conclusions

Google Scholar reveals its most influential papers

1. "Deep Residual Learning for Image Recognition" **(2016)** *Proceedings of the IEEE/CVF Conf. on Computer Vision and Pattern Recognition* 25,256 citations
2. "Deep learning" **(2015)** *Nature* 16,750 citations
3. "Going Deeper with Convolutions" **(2015)** *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition* 14,424 citations
4. "Fully Convolutional Networks for Semantic Segmentation" **(2015)** *Proceedings of the IEEE Conf. on Computer Vision and Pattern Recognition* 10,153 citations
5. "Prevalence of Childhood and Adult Obesity in the United States, 2011-2012" **(2014)** *JAMA* 8,057 citations
6. "Global, regional, and national prevalence of overweight and obesity in children and adults during 1980–2013: a systematic analysis for the Global Burden of Disease Study 2013" **(2014)** *Lancet* 7,371 citations
7. "Observation of Gravitational Waves from a Binary Black Hole Merger" **(2016)** *Physical Review Letters* 6,009 citations

Deep Learning and society expectation



[Deep Learning's 'Permanent Peak' On Gartner's Hype Cycle]

Multi-Task Learning (MTL)

Cuisine: French.

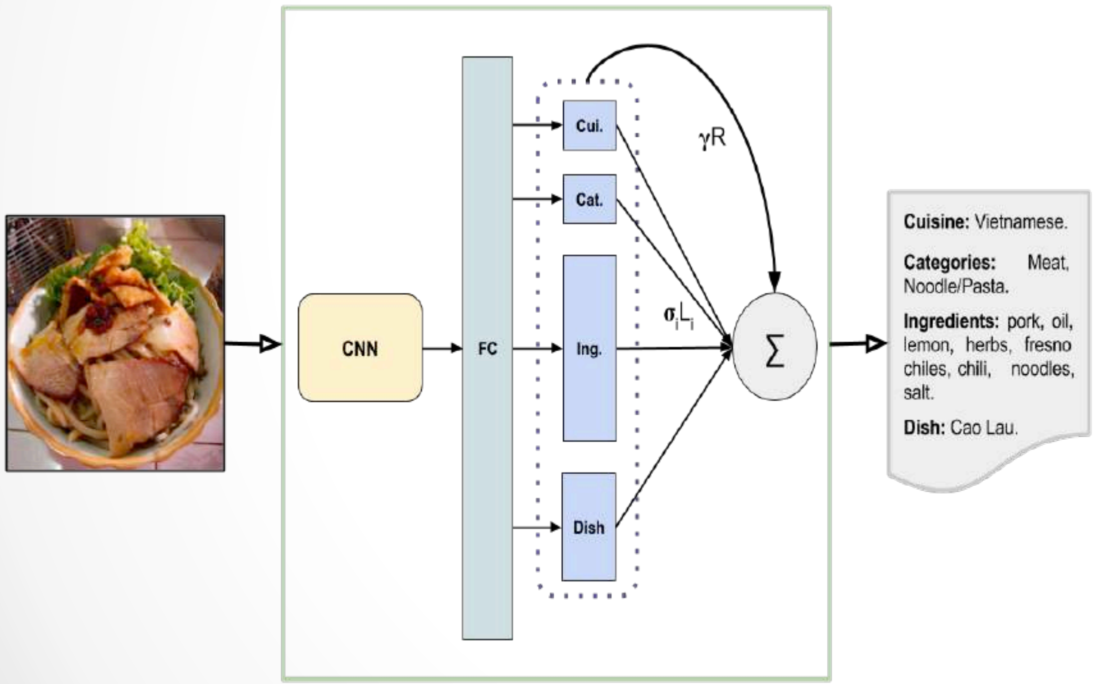
Categories: Meat.

Ingredients: salt, oil, onion, garlic, black pepper, tomato, cloves, parsley, thyme, bay, white wine, clove, duck, fat, mutton.

Dish: Confit de canard.

- Learning **multiple objectives** from a shared representation
  - *<u>Efficiency</u>* and prediction *<u>accuracy</u>*.
- Crucial importance in systems where **long computation** run-time is prohibitive
  - Combining all tasks *<u>reduces computation</u>*.
- Inductive **knowledge transfer**
  - *<u>Generalization</u>* by sharing the domain information between complimentary tasks.

Food Recognition as a MTL



$$L_{total} = \sum_i \omega_i L_i$$

How to define the importance of each task?

- Weighted uniformly the losses.
- Manually tuned the losses.
- Dynamic weighted of the losses.
  - The main task is fixed and weights are learned for each side-task ([1]).
  - Weight the tasks according to the homoscedastic uncertainty ([2]).

[1] X. Yin and X. Liu. Multi-task convolutional neural network for face recognition.
[2] A. Kendall, Y. Gal, and R. Cipolla. Multi-task learning using uncertainty to weigh losses for scene geometry and semantics.


Let’s talk about uncertainty
09:42

Model uncertainty

1. Given a model trained with several pictures of fruits, a user asks the model to decide what is the object using a photo of a chocolate cake.

Who is the guilty for this?

$$S(y_i) = \frac{e^{y_i}}{\sum_j e^{y_j}}$$

Adapted from Gal (2016)

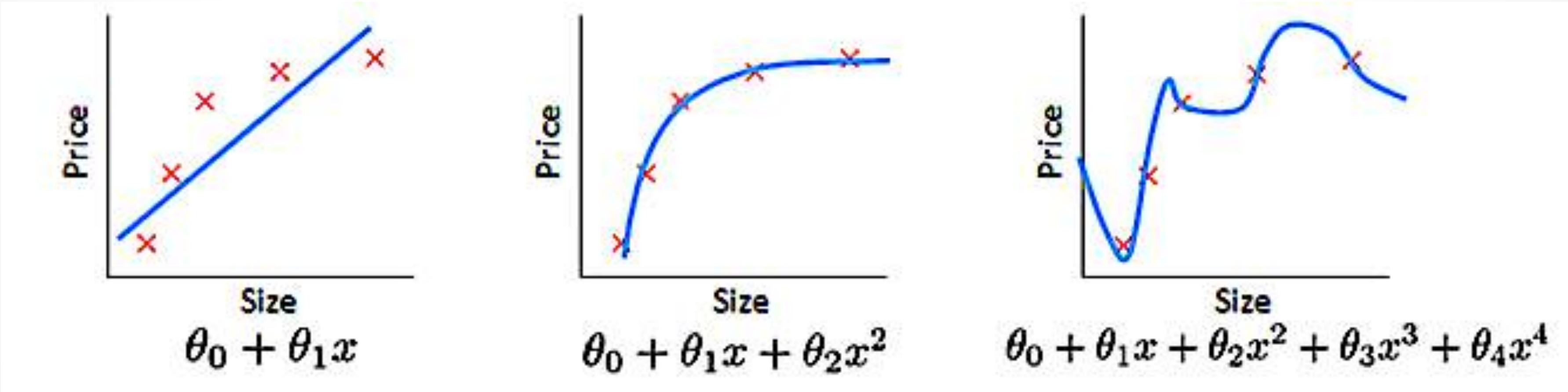
Model uncertainty

2. We have different types of images to classify fruits, where one of the category comes with a lot of clutter/noise/occlusions.

Adapted from Gal (2016)

Model uncertainty

3. What is the best model parameters that best explain a given dataset? What model structure should we use?



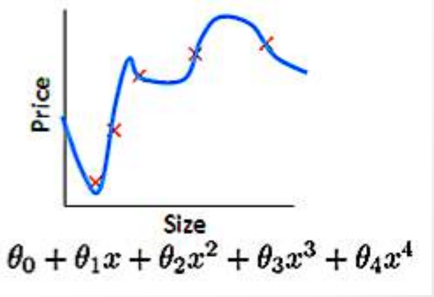
Gal (2016)

Types of uncertainty in Bayesian modeling

Aleatoric – captures the noise inherent in the observations

- heteroscedastic – data-dependent
- homoscedastic – constant for different data points,
  - but can be task-dependent.

- **Epistemic** – model uncertainty
  - Can be explained away given enough data
  - Uncertainty about the model parameters
  - Uncertainty about the model structure

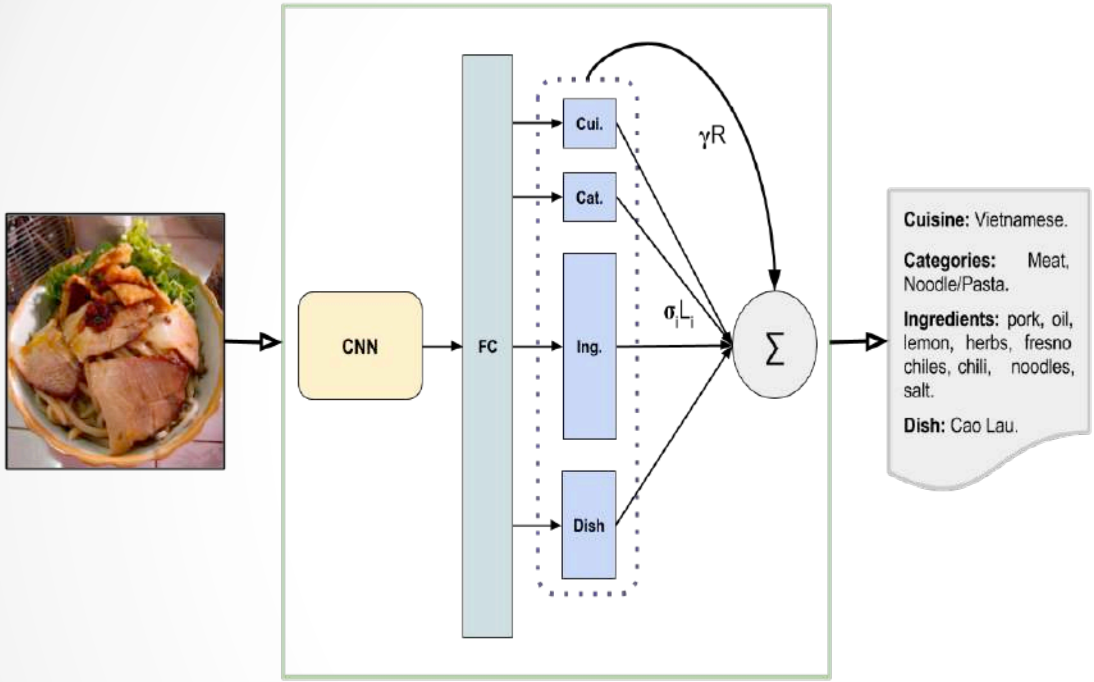


Contents

- The food image problem
- Deep learning for food analysis
- Uncertainty modelling
- Multi-task food learning with aleatoric uncertainty
- Hierarchical food recognition with epistemic uncertainty
- Embedding food ontology
- Conclusions

Food Recognition as a MTL

Aleatoric uncertainty – How to model it?



$$L_{total} = \sum_i \omega_i L_i$$

How to determine the total loss of the MTF?

- Expensive to learn & Affects the performance and the efficiency.

Use aleatoric uncertainty modeling to make the model smarter!

Multi-task uncertainty-based likelihood

In maximum likelihood inference, we maximize the log likelihood of the model:

$$L(W, \sigma, ..., \sigma) = -\log p(y_1, ...y_T | f^W(x))$$

Kendal et.al. (Kendal'2016) showed that:

$$L(W, \sigma, ..., \sigma) = -\log p(y_1, ...y_T | f^W(x)) \approx \sum_{i=1}^{T} (\frac{1}{2\sigma_i^2} L_i(W) + \log \sigma_i^2)$$

- **Proved that the formula can be extended for the binary cross entropy too (multi-label problems).**

Eduardo Aguilar, Marc Bolaños, Petia Radeva: **Regularized uncertainty-based multi-task learning model for food analysis.** J. Visual Communication and Image Representation 60: 360-370 (2019)

Validation

FoodImageNet

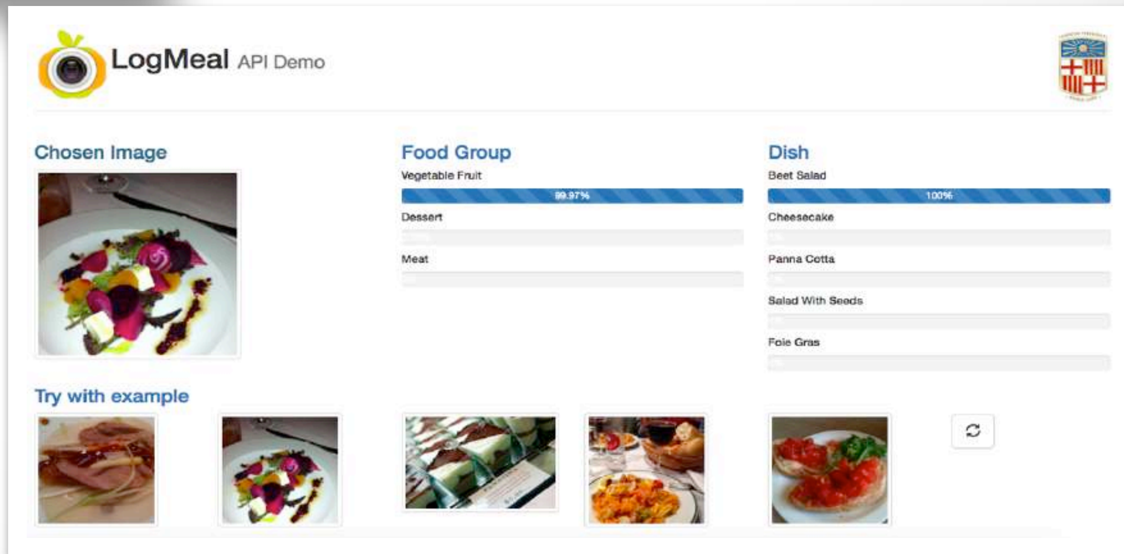
- Food – 550 dishes, 11 categories, 11 cuisines
- Ingredients – 65
- Drinks – 40

In total:
more than
550.000 images

Eduardo Aguilar, Marc Bolaños, Petia Radeva: **Regularized uncertainty-based multi-task learning model for food analysis.** J. Visual Communication and Image Representation 60: 360-370 (2019)

Food ingredients recognition

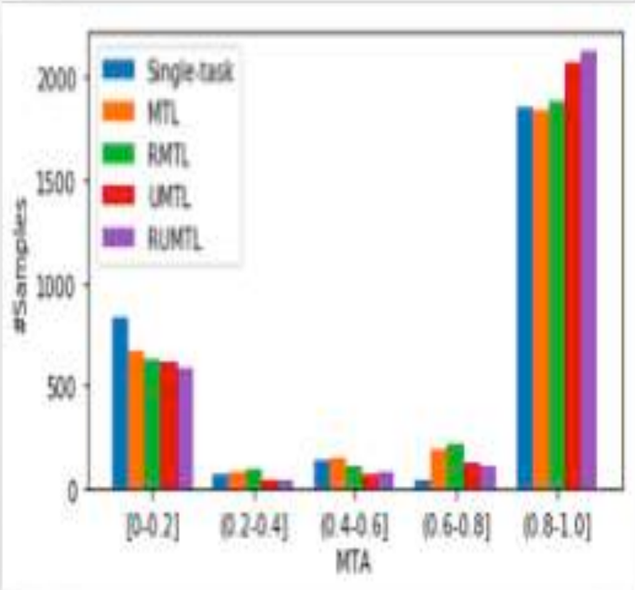
[Food category and class recognition]

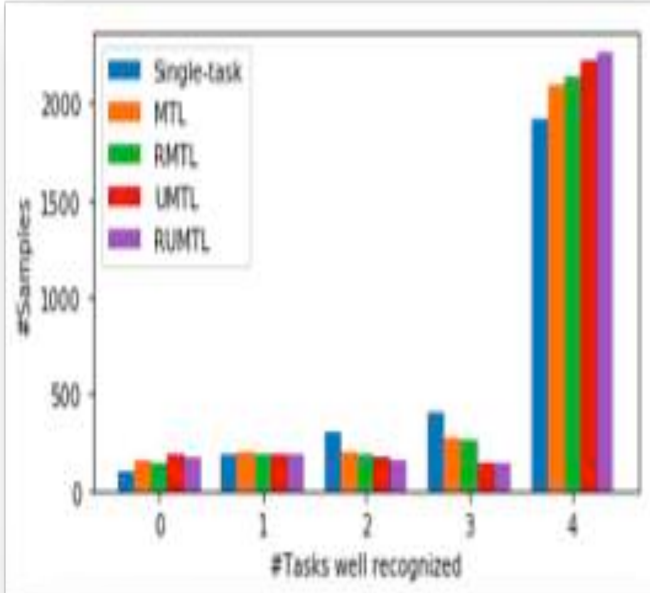


Food ingredients recognition

	Dish	Cuisine	Categories			Ingredients			
	Acc	*Acc*	F_1	*Pre*	*Rec*	F_1	*Pre*	*Rec*	*MTA*
Single-task	0.8334	0.8649	0.8709	0.8944	0.8485	**0.8992**	0.9143	0.8846	0.6713
MTL	0.8303	**0.8958**	0.8811	0.9042	0.8592	0.8780	0.8972	0.8596	0.6927
RMTL	0.8351	0.8917	0.8834	0.8789	0.8880	0.8809	0.8613	0.9014	0.7061
UMTL	0.8221	0.8944	0.8925	0.9067	0.8788	0.8943	0.9095	0.8795	0.7478
RUMTL	**0.8358**	0.8934	**0.8944**	0.9041	0.8848	0.8988	0.9084	0.8893	**0.7600**

Multi-task Accuracy: encourage errors to concentrate on the same data.





Contents

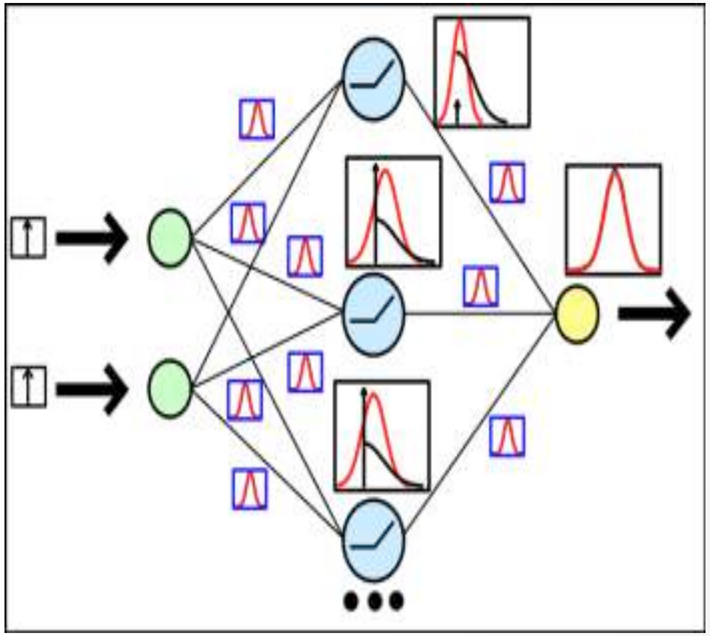
- The food image problem
- Deep learning for food analysis
- Uncertainty modelling
- Multi-task food learning with aleatoric uncertainty
- Hierarchical food recognition with epistemic uncertainty
- Embedding food ontology
- Conclusions

Bayesian neural networks

Instead of learning the model's weights,
learn a distribution over the weights

- => estimate uncertainty over the weights.

- So how do we do that?

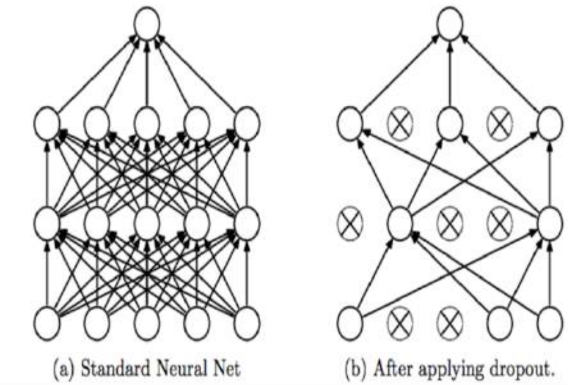


[Financial forecasting with probabilistic programming and Pyro]

How to estimate the Epistemic Uncertainty?

Gal and Ghahramani showed that dropout at inference time gives an uncertainty estimator:

1. Infer y|x multiple times, each time sample a different set of nodes to drop out.
2. Average the predictions to get the final prediction E(y|x).
3. Calculate the sample variance of the predictions.



How to estimate the Epistemic Uncertainty?

The Epistemic Uncertainty (EU) can be expressed as follows:

where

$$EU(x_t) = -\sum_{c=1}^{C} \overline{p(y_c = \hat{y}_c | x_t)} \ln(\overline{p(y_c = \hat{y}_c | x_t)}),$$

K Monte Carlo dropout simulations

$$\overline{p(y_c = \hat{y}_c | x)} = \frac{1}{K} \sum_{k=1}^{K} p(y_c^k = \hat{y_c^k} | x).$$

How many dishes there are all over the world?

WIKIPEDIA
The Free Encyclopedia

More than
100.000 basic
foods

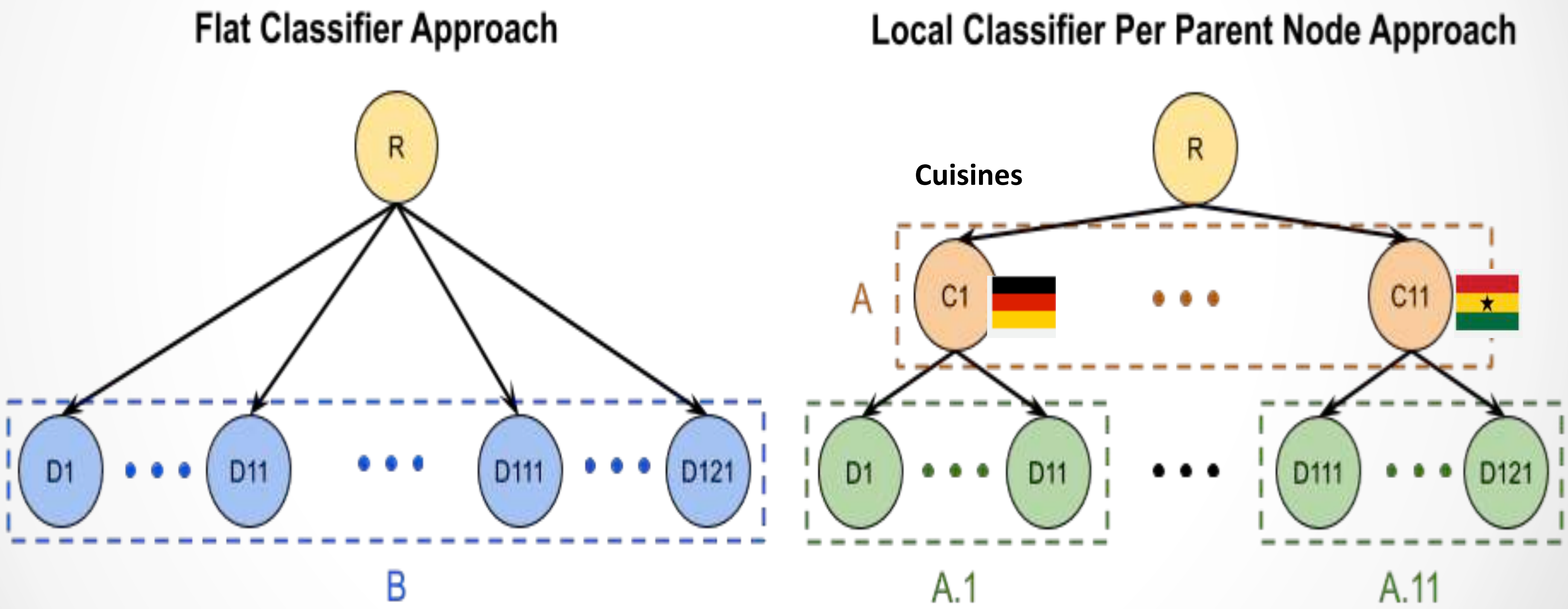
Imagine

- When you visit Mexico,

what is the probability to eat a food from Norway?

Let’s organize classes in meta-classes

Let’s organize classes in meta-classes

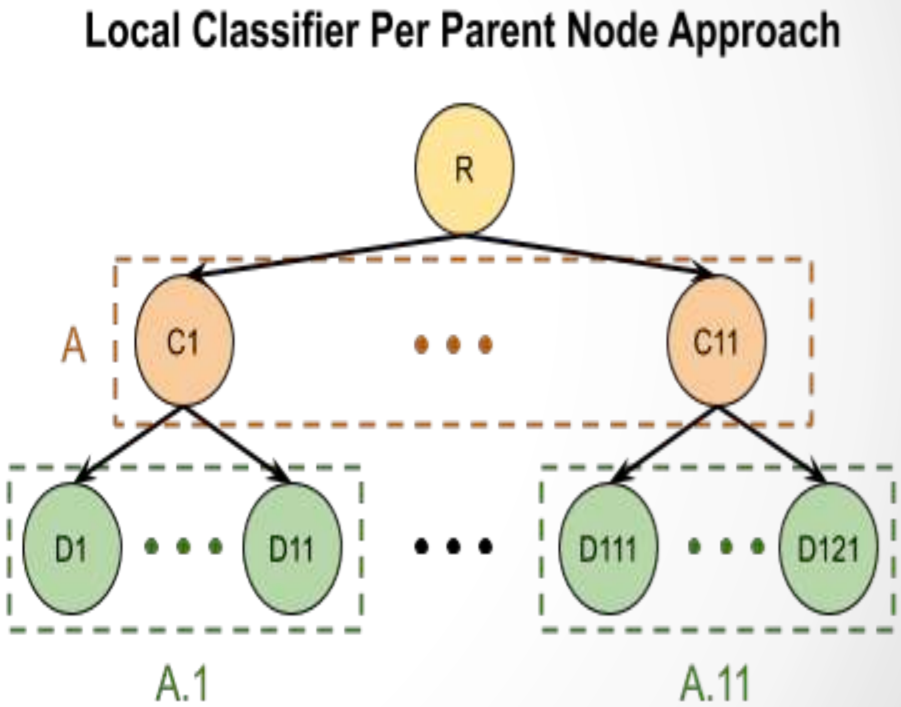


Dishes

Dishes

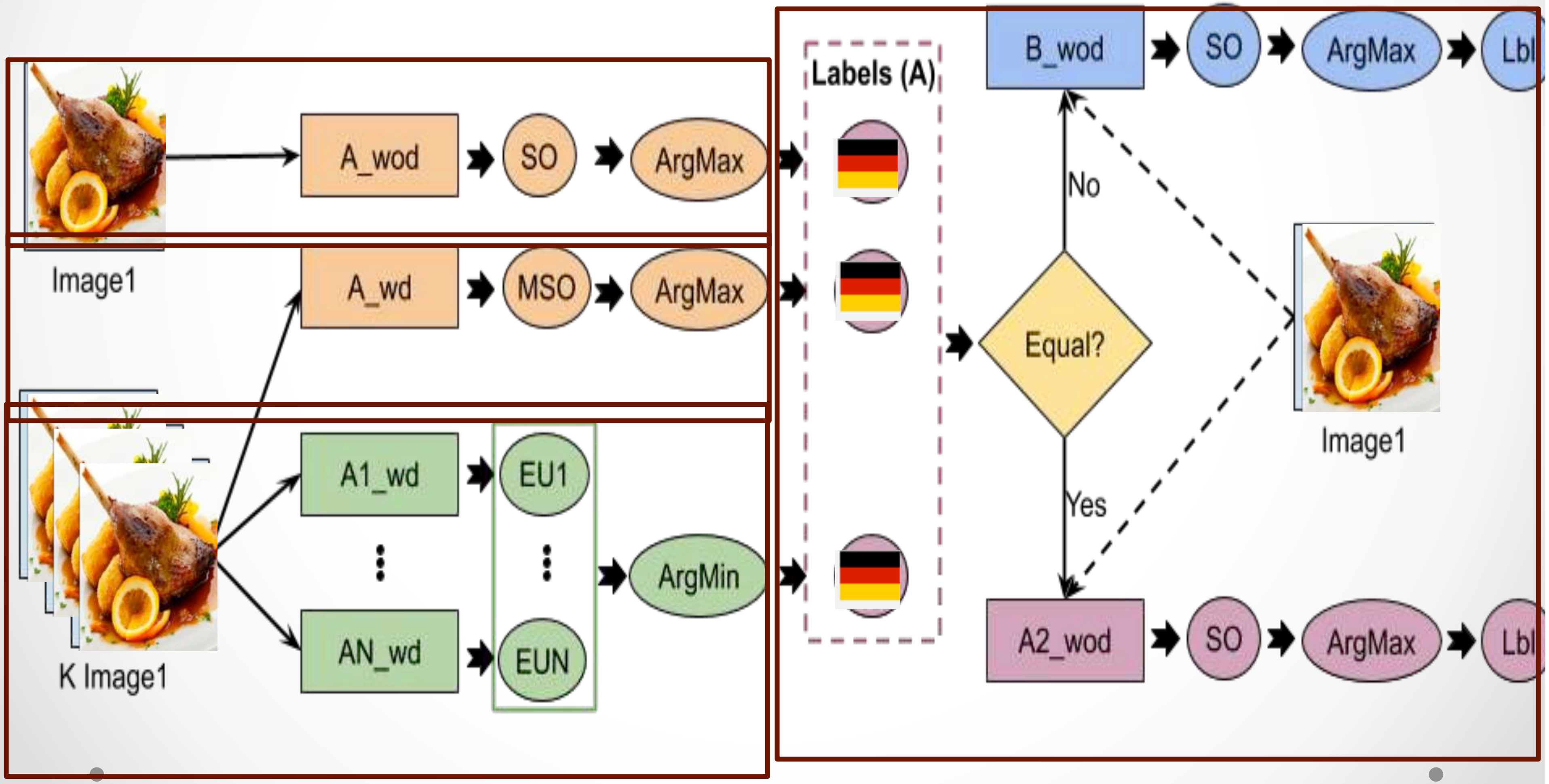
But …. Hierarchical classifiers have a big problem

Error propagation



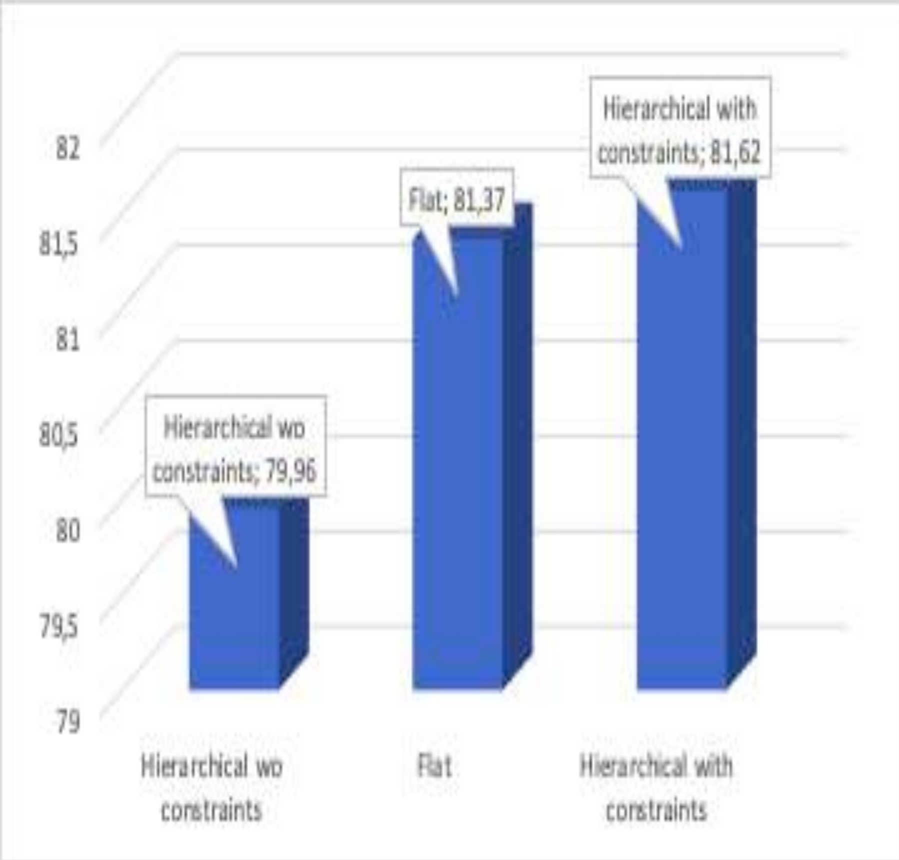
Hypothesis: use uncertainty to decide if a LPN should be used

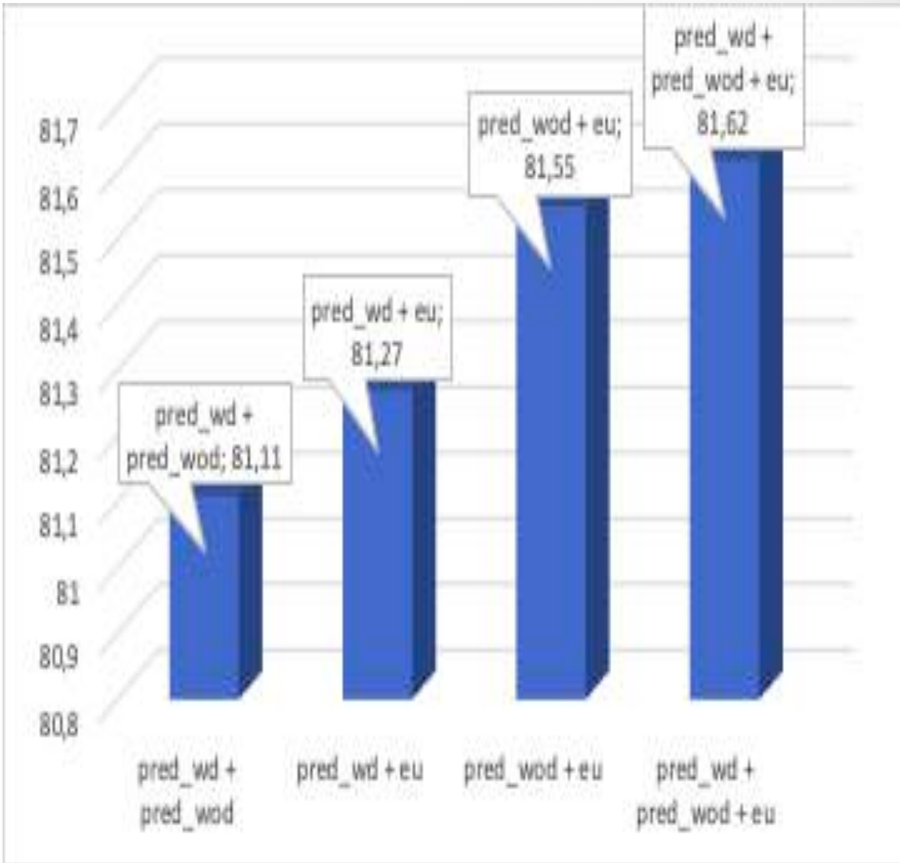
Proposed Method



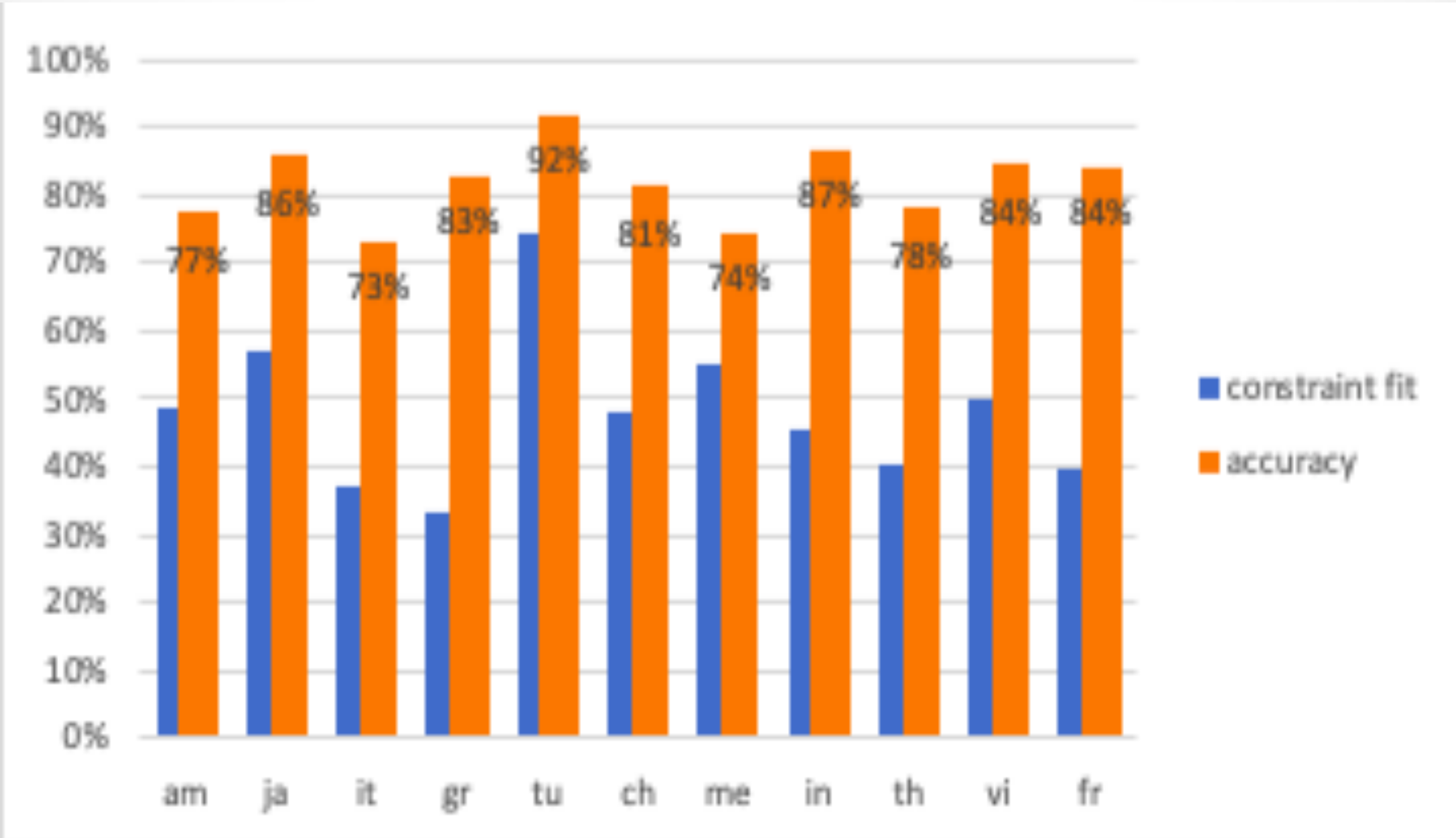
Validation

Ablation study





Rate of images fulfilling the criterion per cuisines and their accuracy



Results - Samples of the Smallest and Largest EU within the same class of Dish

Caesar Salad

Ravioli

Steak

Tacos

Contents

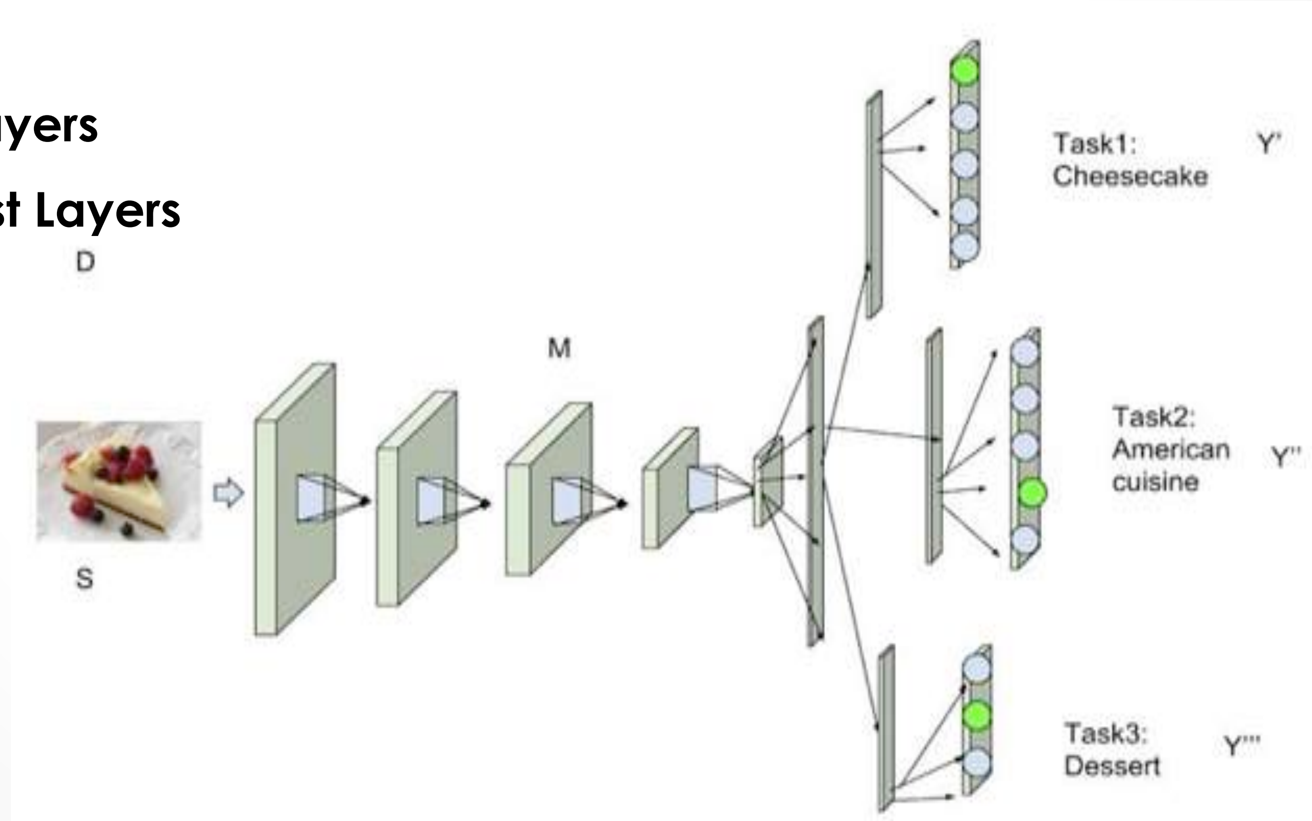
- The food image problem
- Deep learning for food analysis
- Uncertainty modelling
- Multi-task food learning with aleatoric uncertainty
- Hierarchical food recognition with epistemic uncertainty
- Embedding food ontology
- Conclusions

Multi-Task Learning Model with Ontology

Multi-Task Learning Model

Shared Base Layers

Specialized Last Layers



Motivation

Hypothesis **Dataset with Multiple Task Labels**

Egg´s Benedict

Dish

Ingredients

Image source: Recipes5k

Eggs

Parsley

Toast bread

Bacon

Hollandaise sauce

Butter

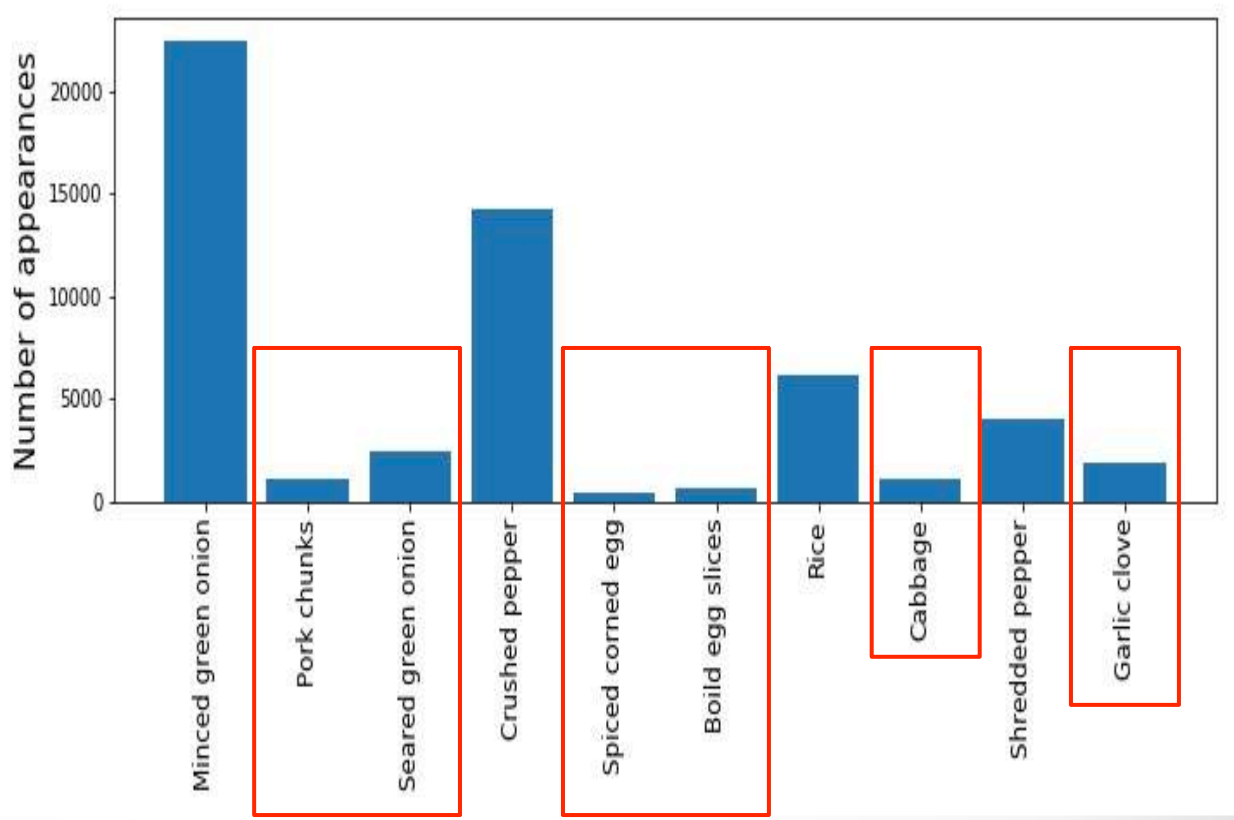
Motivation

Food Analysis Problems

Less Frequent Ingredients in dataset

Difficult to detect

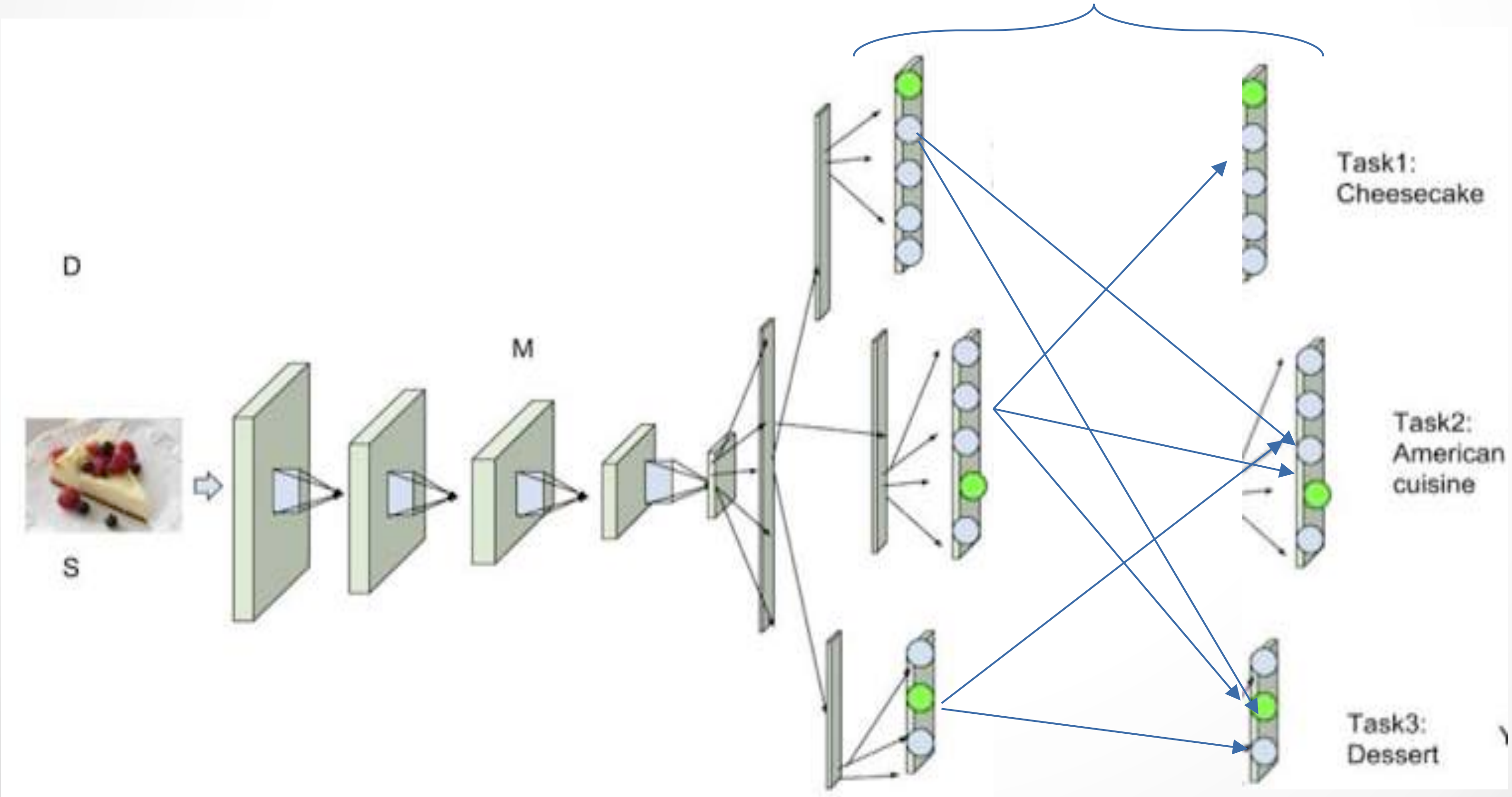
Low Precision



Multi-Task Learning Model with Ontology

Multi-Task Learning Model

Relations between tasks



Multi-Task Learning Model with Ontology

How to convert it into a Layer?

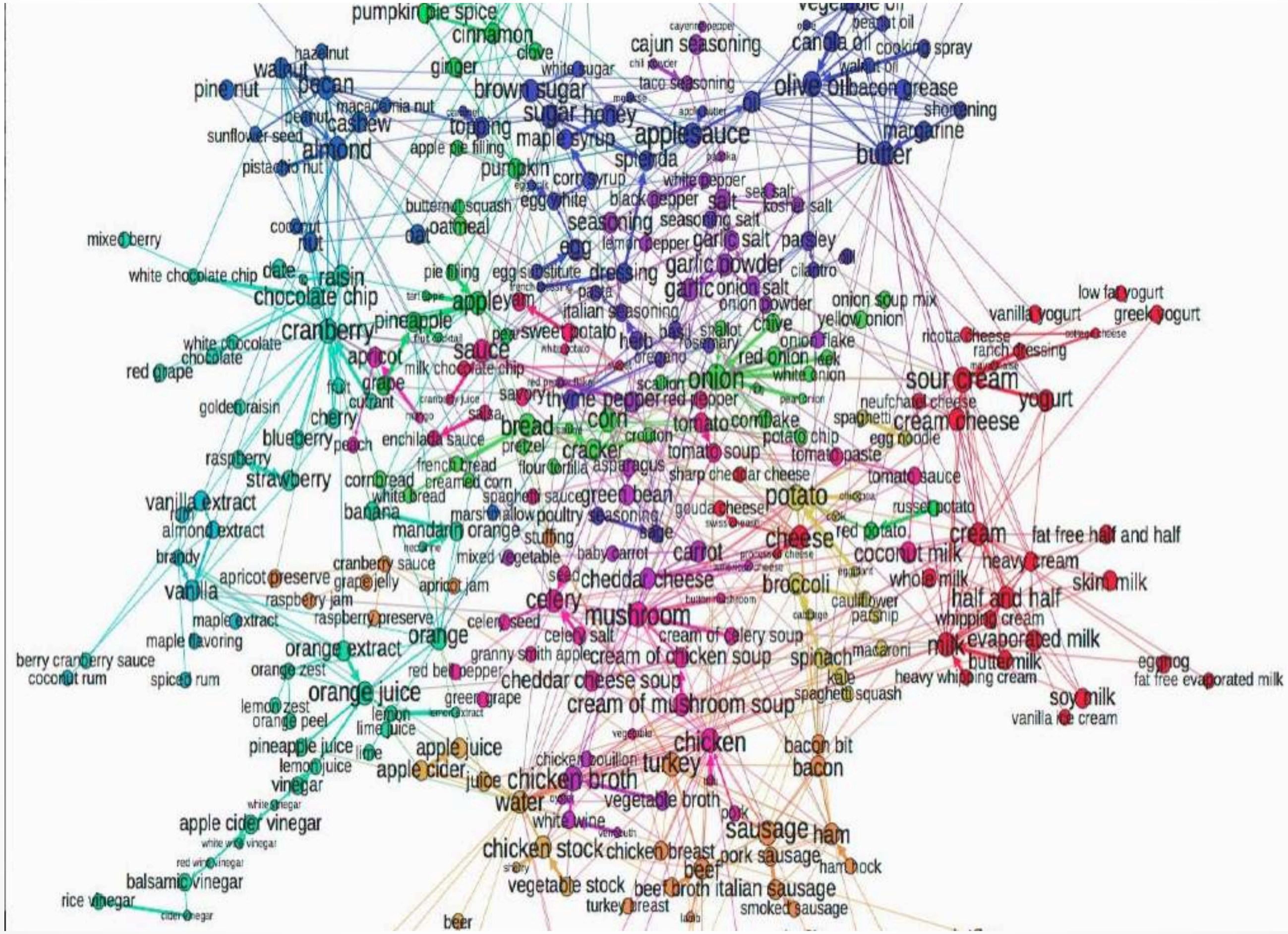


Image source: ladamic

Multi-Task Learning Model with Ontology

Ontology Layer

⇩

Relations between classes of same and different tasks

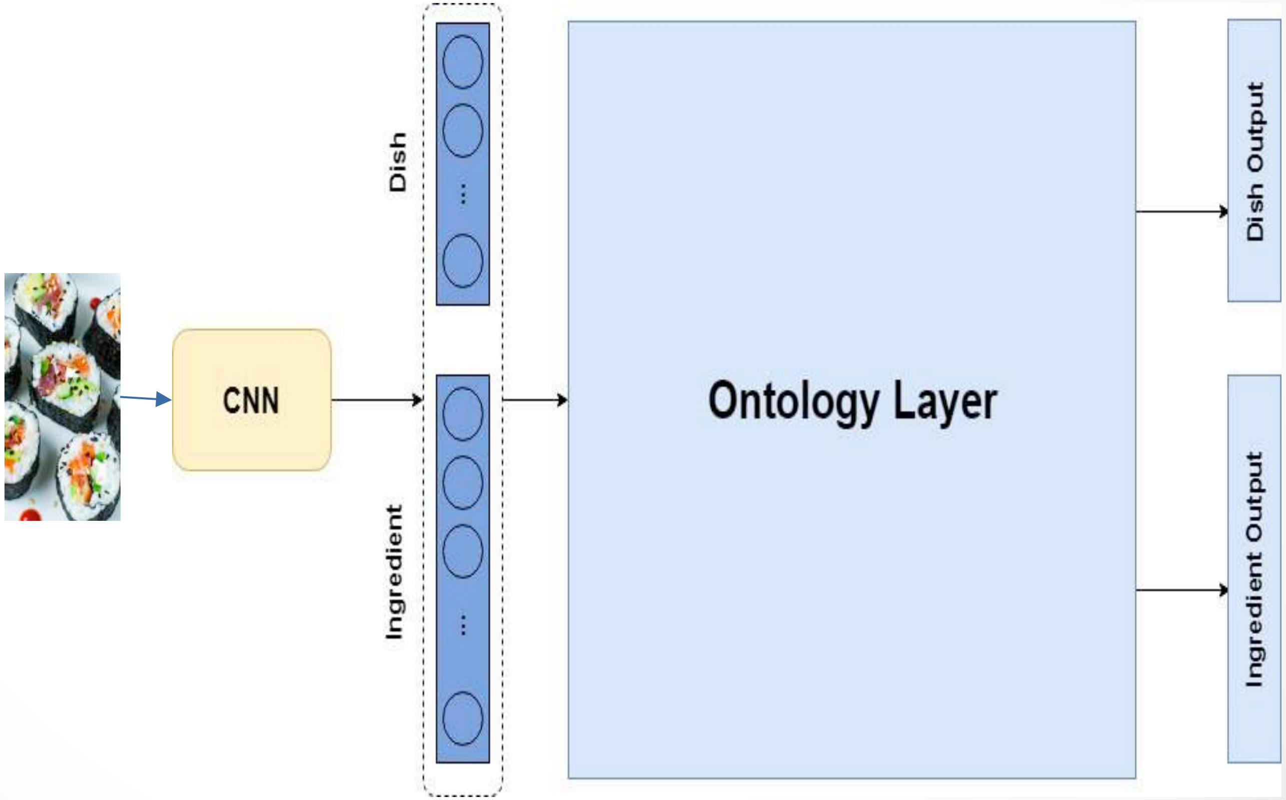


Image source: Applying Deep Learning for Food Image Analysis

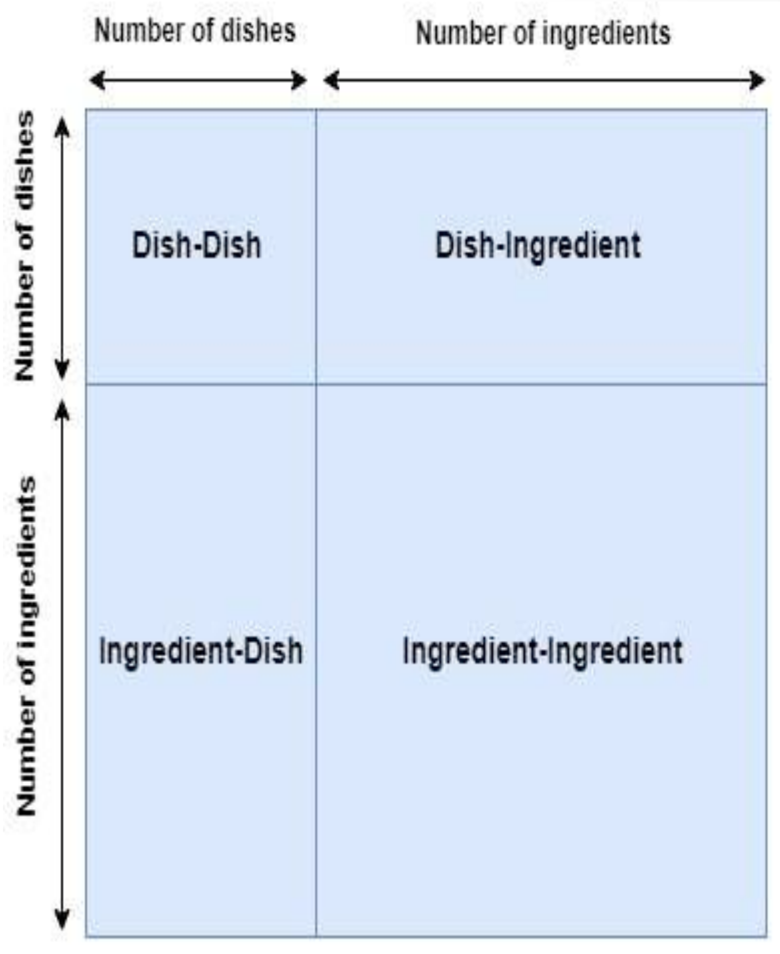
Multi-Task Learning Model with Ontology

Matrix

#elements x #elements

Relations

- Dish-Dish
- Dish-Ingredient
- Ingredient-Dish
- Ingredient-Ingredient



Multi-Task Learning Model with Ontology

Ontology

- Element values

Ontology made of probabilities and “negative probabilities”

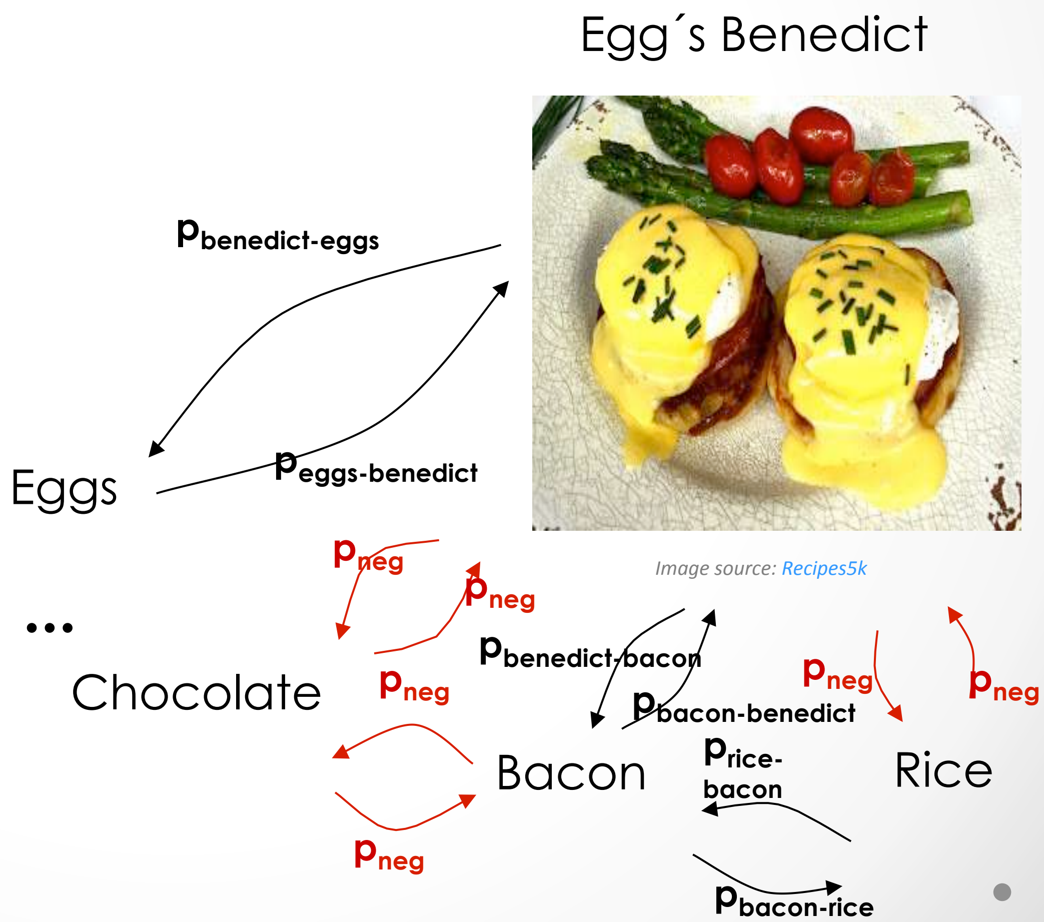
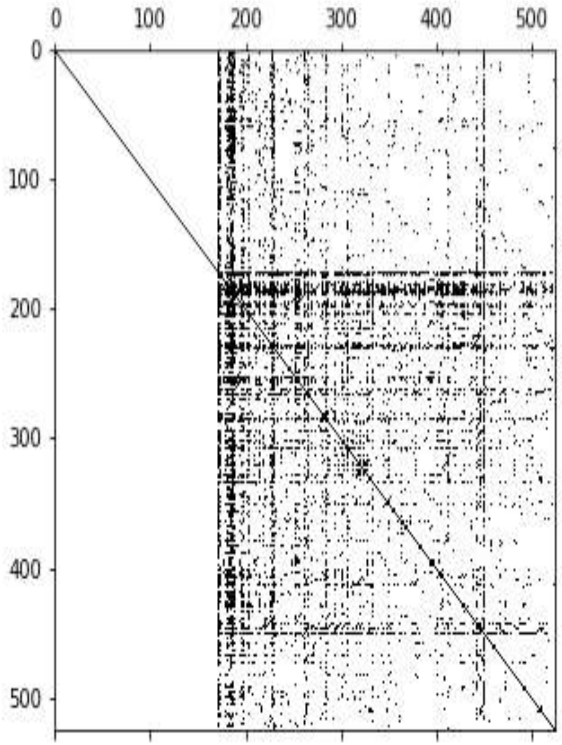


Image source: Recipes5k

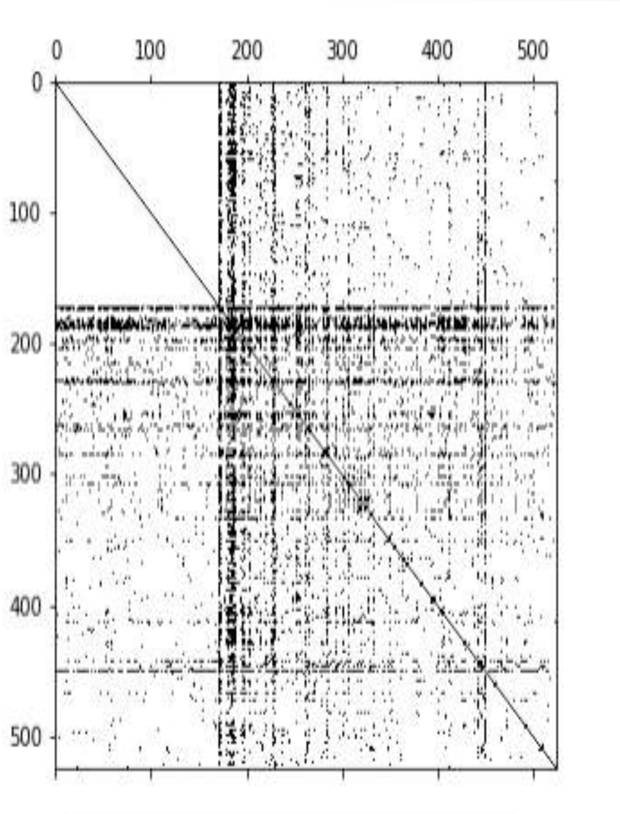
Multi-Task Learning Model with Ontology

Ontology

- Structure



Dish-Ingredient
Ingredient-Ingredient

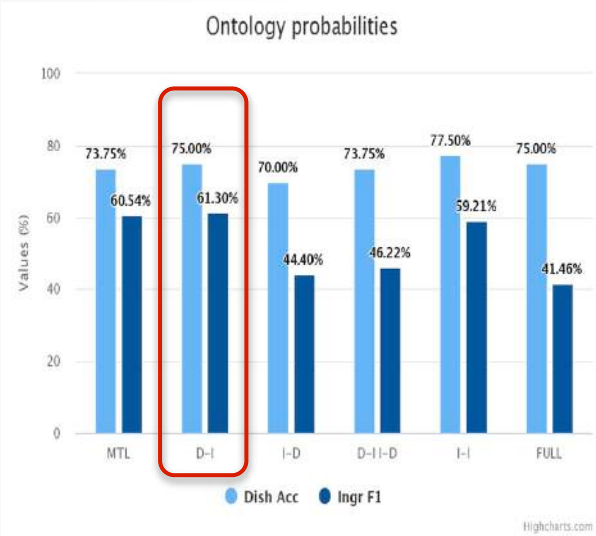


Full

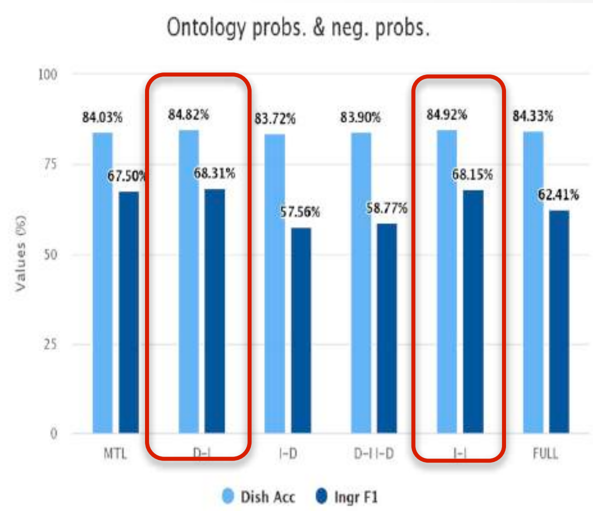
Validation

Experimental Results

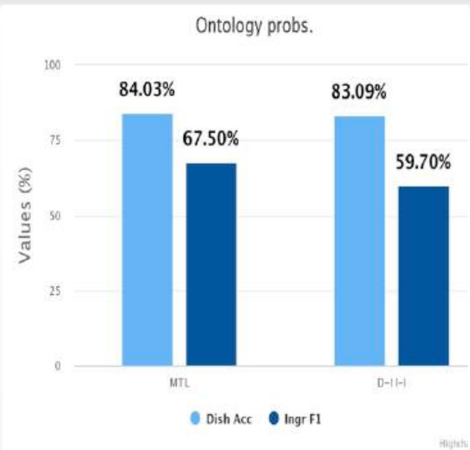
Recipes5k Results

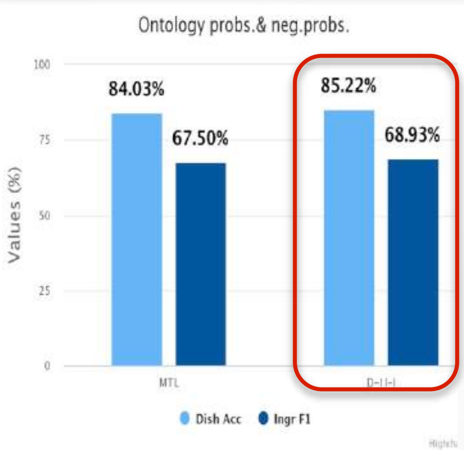


VireoFood-172 Results



Dish-Ingr
Ingr-Ingr



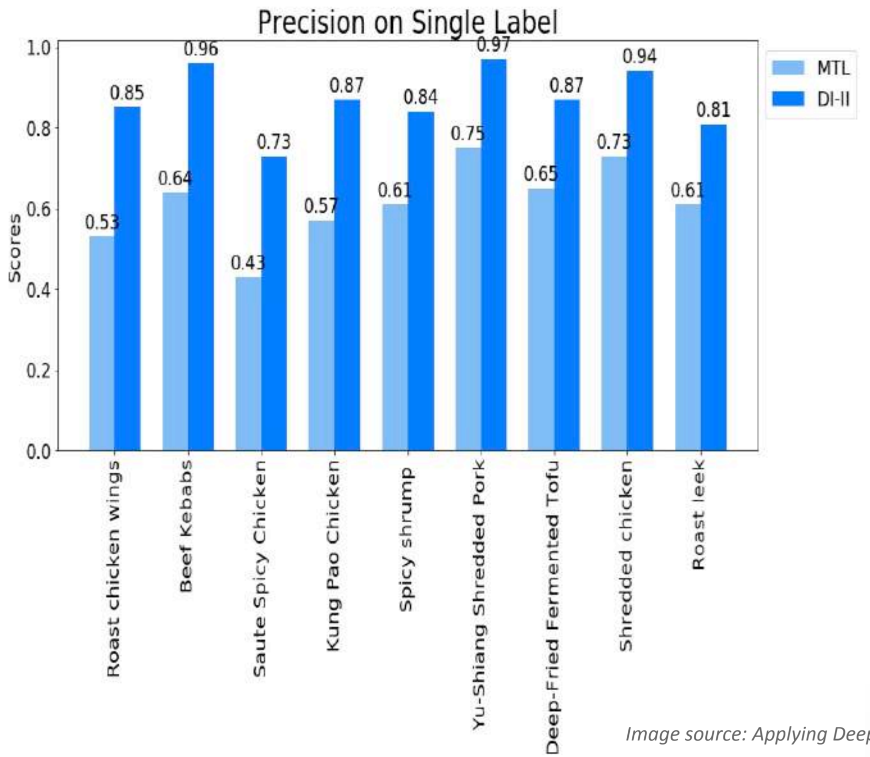


DI-II

Best Performance

Experimental Results

MTL vs D-I I-I Ontology Model



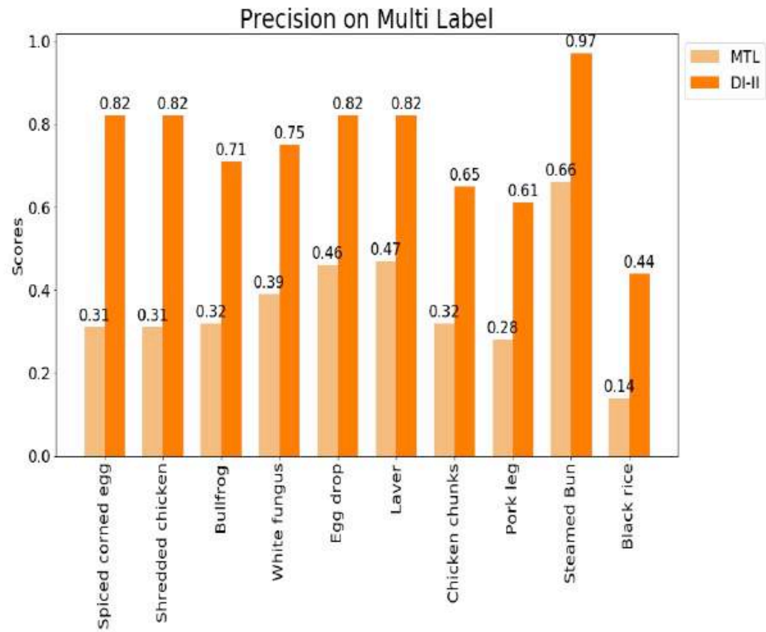


Image source: Applying Deep Learning for Food Image Analysis

Conclusions

- Food image world brings us huge amount of data and Computer Vision questions

- Transfer learning and its subproblems (multi-task learning) open new opportunities

- Uncertainty modeling is a hot topic with many open questions and challenges!
  - Exclusivity relation between elements helps to the classification

  - Epistemic uncertainty

    New method for robust hierarchical classifiers..
    A good cue to improve recognition scalability.
    Epistemic uncertainty useful beyond the confidence of the model.
  - Aleatoric uncertainty

    Allows to weight different tasks according to uncertainty

  - For first time a food ontology is integrated into an end-to-end model

A huge impact of food analysis is expected from point of view of:
- Science, but also
- Real world applications, specially important for the society.

What is the next?

Using GANs to augment data


Thank you!