Uncertainty-Aware Food Recognition by Deep Learning

Petia Radeva,

Collaboration with:
Eduardo Aguilar, Marc Bolaños

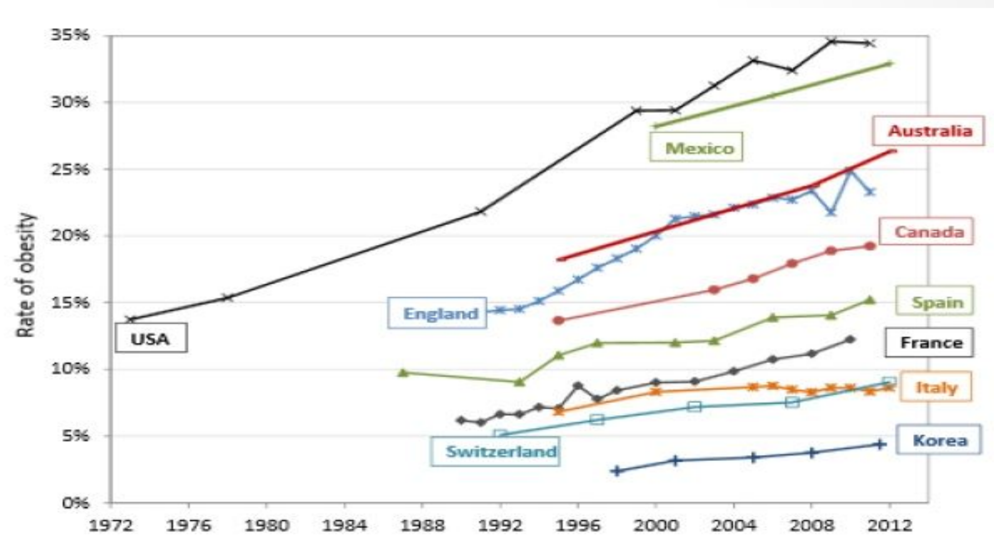
University of Barcelona &
Computer Vision Center

radevap@gmail.com

The Diabetes pandemy

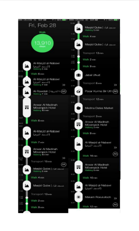
Diabetic people need to follow a strict record of their meals!

Chronic disease statistics



What are we missing in health applications?

- Today, automatically measuring physical activity is not a problem.
- But what about food and nutrition?



Identificador del voluntario: **Fecha del examen:**

Voluntario | Visita | Día / Mes / Año

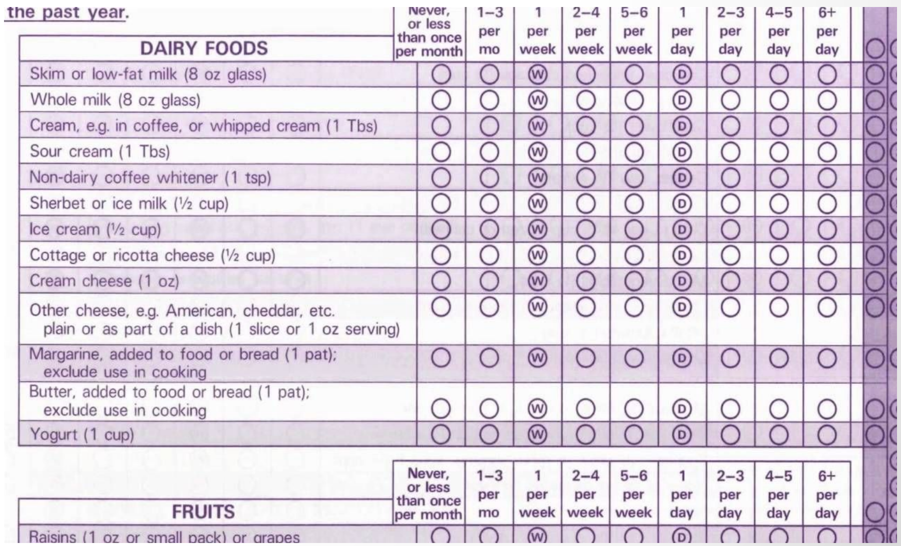
Visita: anotar el número de visita

00. Inclusión/Exclusión / 01. Visita 1 / 02. Visita 2 / 03. Vista 3

FORMULARIO ENCUESTA ALIMENTARIA:

REGISTRO 3 DIAS

Comida	Hora	Alimento o preparación	Ingredientes	Cantidad total (gr)*
Desayuno				
Media mañana				
Comida				



the past year.

DAIRY FOODS	Never, or less than once per month	1–3 per mo	1 per week	2–4 per week	5–6 per week	1 per day	2–3 per day	4–5 per day	6+ per day
Skim or low-fat milk (8 oz glass)	◯	◯	Ⓦ	◯	◯	Ⓓ	◯	◯	◯
Whole milk (8 oz glass)	◯	◯	Ⓦ	◯	◯	Ⓓ	◯	◯	◯
Cream, e.g. in coffee, or whipped cream (1 Tbs)	◯	◯	Ⓦ	◯	◯	Ⓓ	◯	◯	◯
Sour cream (1 Tbs)	◯	◯	Ⓦ	◯	◯	Ⓓ	◯	◯	◯
Non-dairy coffee whitener (1 tsp)	◯	◯	Ⓦ	◯	◯	Ⓓ	◯	◯	◯
Sherbet or ice milk (½ cup)	◯	◯	Ⓦ	◯	◯	Ⓓ	◯	◯	◯
Ice cream (½ cup)	◯	◯	Ⓦ	◯	◯	Ⓓ	◯	◯	◯
Cottage or ricotta cheese (½ cup)	◯	◯	Ⓦ	◯	◯	Ⓓ	◯	◯	◯
Cream cheese (1 oz)	◯	◯	Ⓦ	◯	◯	Ⓓ	◯	◯	◯
Other cheese, e.g. American, cheddar, etc. plain or as part of a dish (1 slice or 1 oz serving)	◯	◯	Ⓦ	◯	◯	Ⓓ	◯	◯	◯
Margarine, added to food or bread (1 pat); exclude use in cooking	◯	◯	Ⓦ	◯	◯	Ⓓ	◯	◯	◯
Butter, added to food or bread (1 pat); exclude use in cooking	◯	◯	Ⓦ	◯	◯	Ⓓ	◯	◯	◯
Yogurt (1 cup)	◯	◯	Ⓦ	◯	◯	Ⓓ	◯	◯	◯

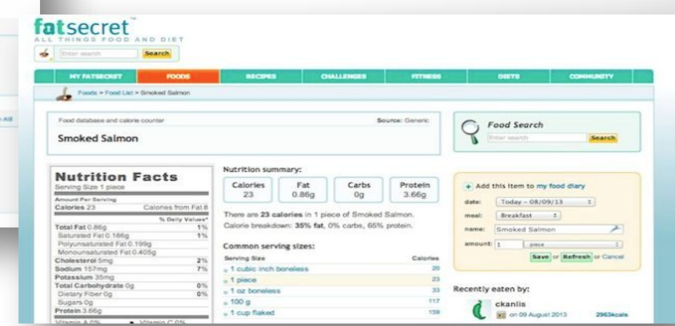
FRUITS	Never, or less than once per month	1–3 per mo	1 per week	2–4 per week	5–6 per week	1 per day	2–3 per day	4–5 per day	6+ per day
Raisins (1 oz or small pack) or grapes	◯	◯	Ⓦ	◯	◯	Ⓓ	◯	◯	◯

What are we missing in health applications?

- But what about food and nutrition?
  - **State of the art:** Nutritional health apps are based on manual food diaries.

Cronometer

Fatsecret



Sparkpeople

LoseIt!

MyFitnessPal

How is today the food intake annotated?

Identificador del voluntario: **Fecha del examen:**

Voluntario Visita Día / Mes / Año

Visita: anotar el número de visita

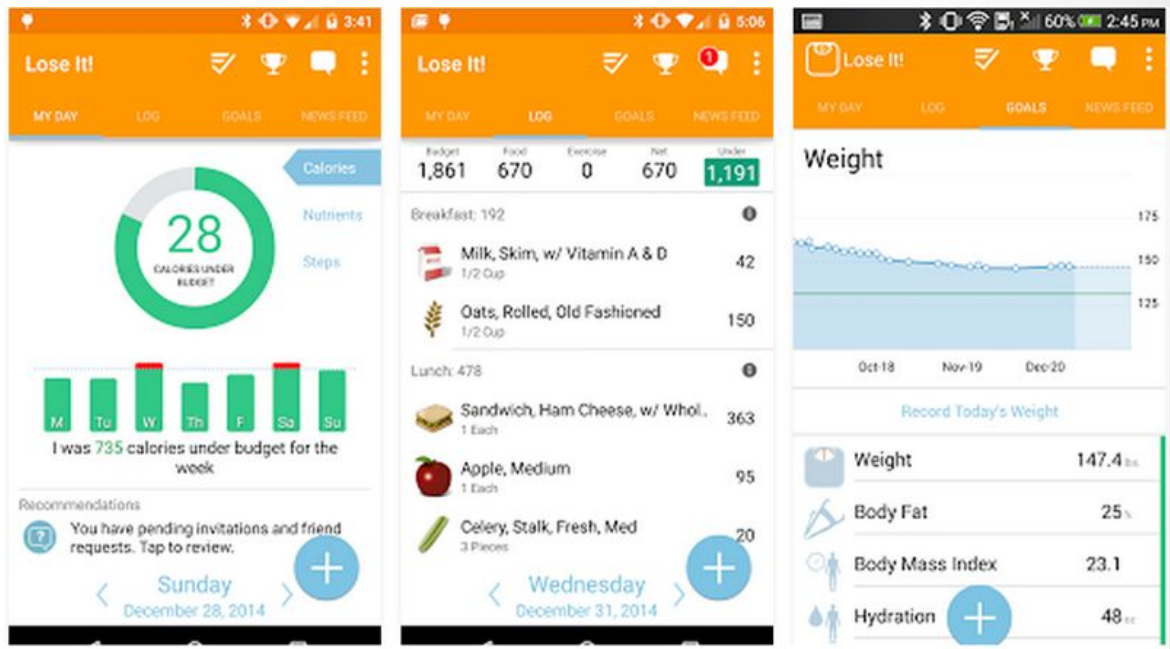
00. Inclusión/Exclusión / 01. Visita 1 / 02. Visita 2 / 03. Vista 3

FORMULARIO ENCUESTA ALIMENTARIA:

REGISTRO 3 DIAS

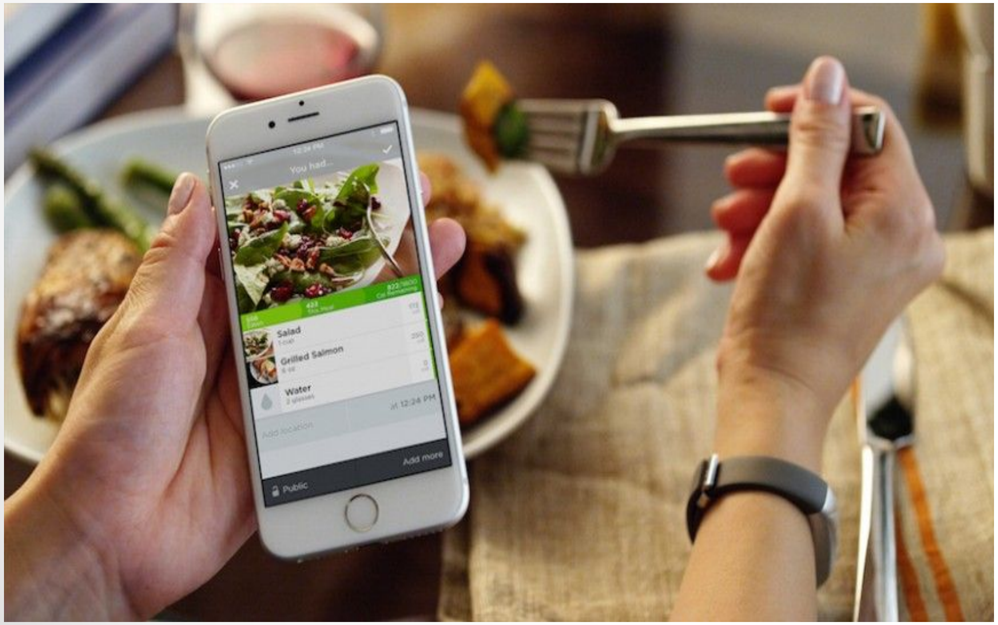
Comida	Hora	Alimento o preparación	Ingredientes	Cantidad total (gr)*
Desayuno				
Media mañana				
Comida				

24 hours dietary recall



What we propose about it?

Automatic visual food recognition tools for dietary assessment.



What about automatic food recognition?


NVIDIA DEVELOPER

NEWS CENTER News Research Events

Comments 4 Shares

Lose It!'s New Food Recognition App Counts Calories

October 4, 2016

The diet app LoseIt! released a new deep learning feature called Snap It that lets users take photos of their food and then it automatically logs the calorie count and nutritional information.

Using the NVIDIA DIGITS deep learning training system on four TITAN X GPUs, the company trained their network on a vast database of 230,000 food images and more than 4 billion foods logged by Lose It! users since 2008.

"The more people use this, the more it improves," said Edward W. Lowe, data scientist at Lose It. "The goal is to get the accuracy high enough in six months so it won't even need to ask you for validation."

How many food categories there are?

Today we are speaking about 200.000 food categories, 8000 basic food (Wikipedia).

Is it possible?

https://techcrunch.com/2016/09/29/lose-it-launches-snap-it-to-let-users-count-calories-in-food-photos/

Why is the food recognition a challenge?

Difficulties

Huge intra-class variations

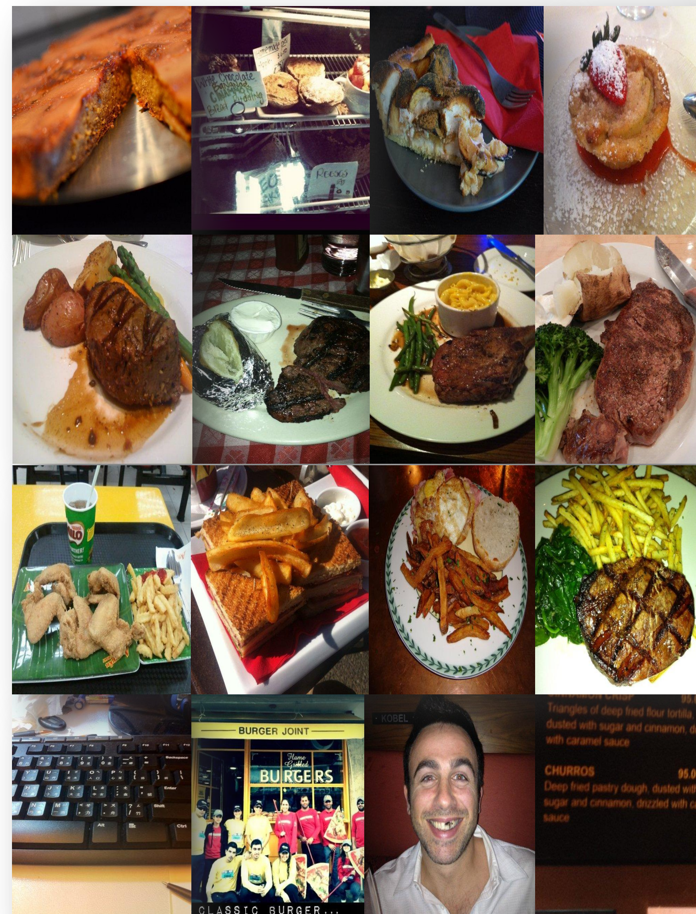
Ambiguous definition

Inter-class similarities

Mixed items

Need of huge datasets

Bad Labeled



What to do when you have a really complicate problem?

Any powerful tools for data processing of large amount of data?

Google Scholar reveals its most influential papers


nature International weekly journal of science

Deep learning

Yann LeCun, Yoshua Bengio & Geoffrey Hinton

Affiliations | Corresponding author

Nature **521**, 436–444 (28 May 2015) | doi:10.1038/nature14539

Received 25 February 2015 | Accepted 01 May 2015 | Published online 27 May 2015

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 applications

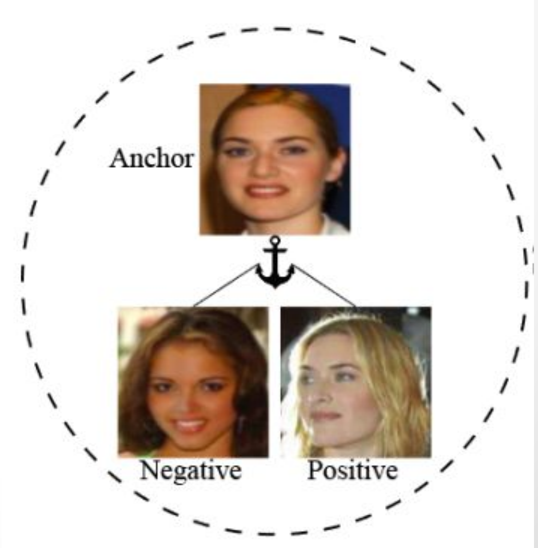
MIT Technology Review

Climate Change Sep 19

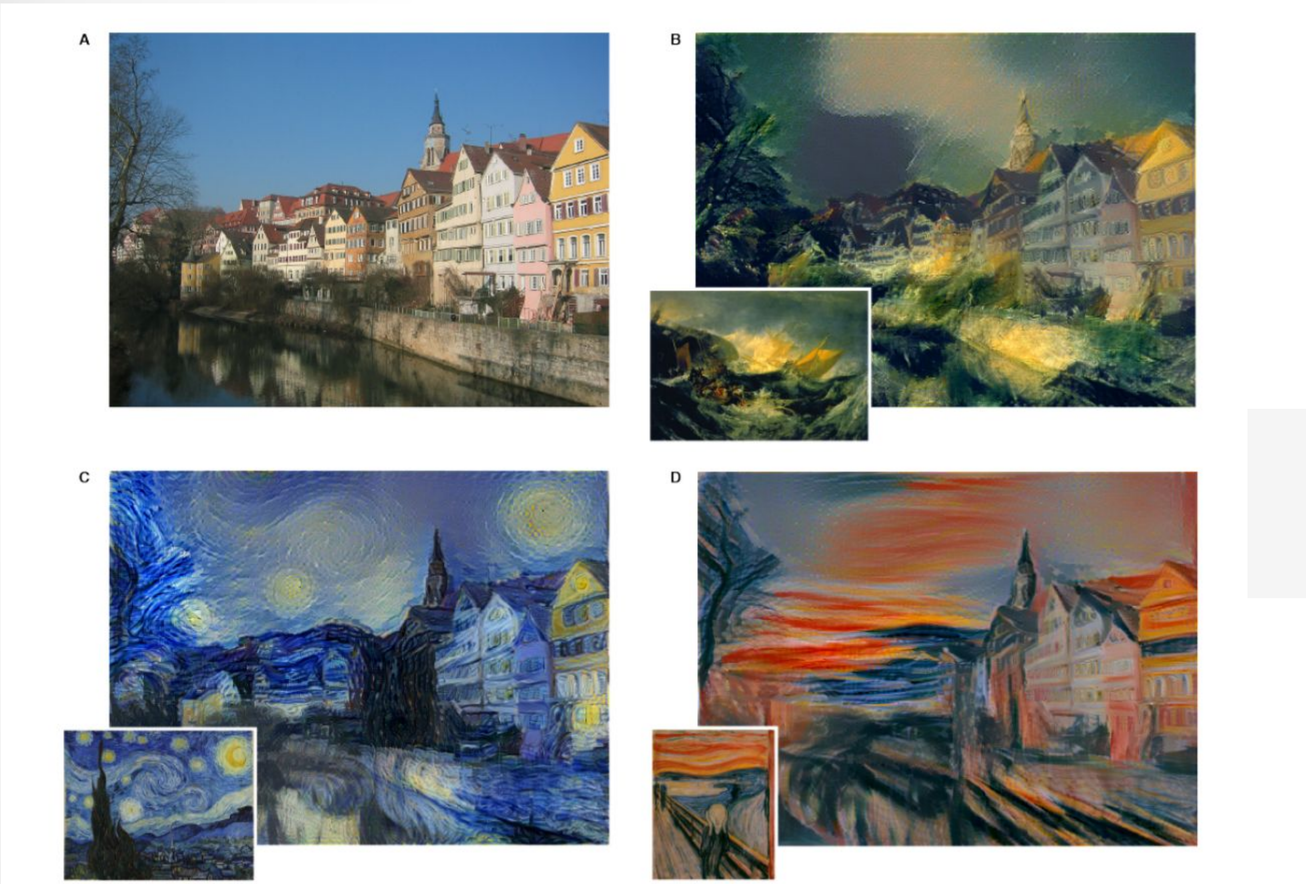
Amazon just pledged to hit net zero climate emissions by 2040

Neural Networks beat humans in:

- object recognition,
- lip reading,
- high-end surveillance,
- facial recognition,
- object-based searches,
- license plate readers,
- traffic violations detection,
- breast tomosynthesis diagnosis,
- etc., etc.



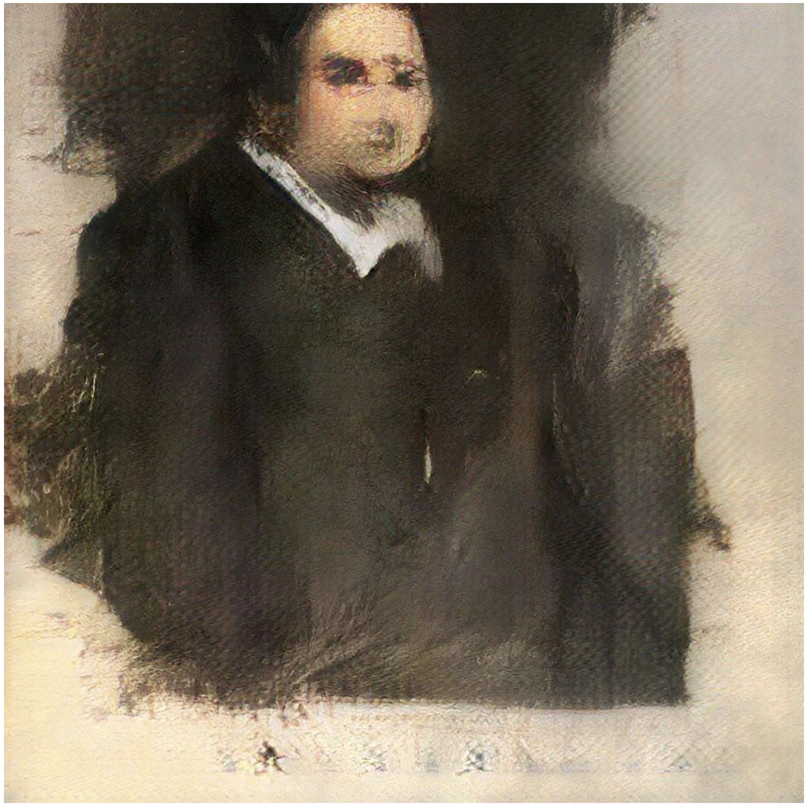
Neural Style Transfer



PRISMA

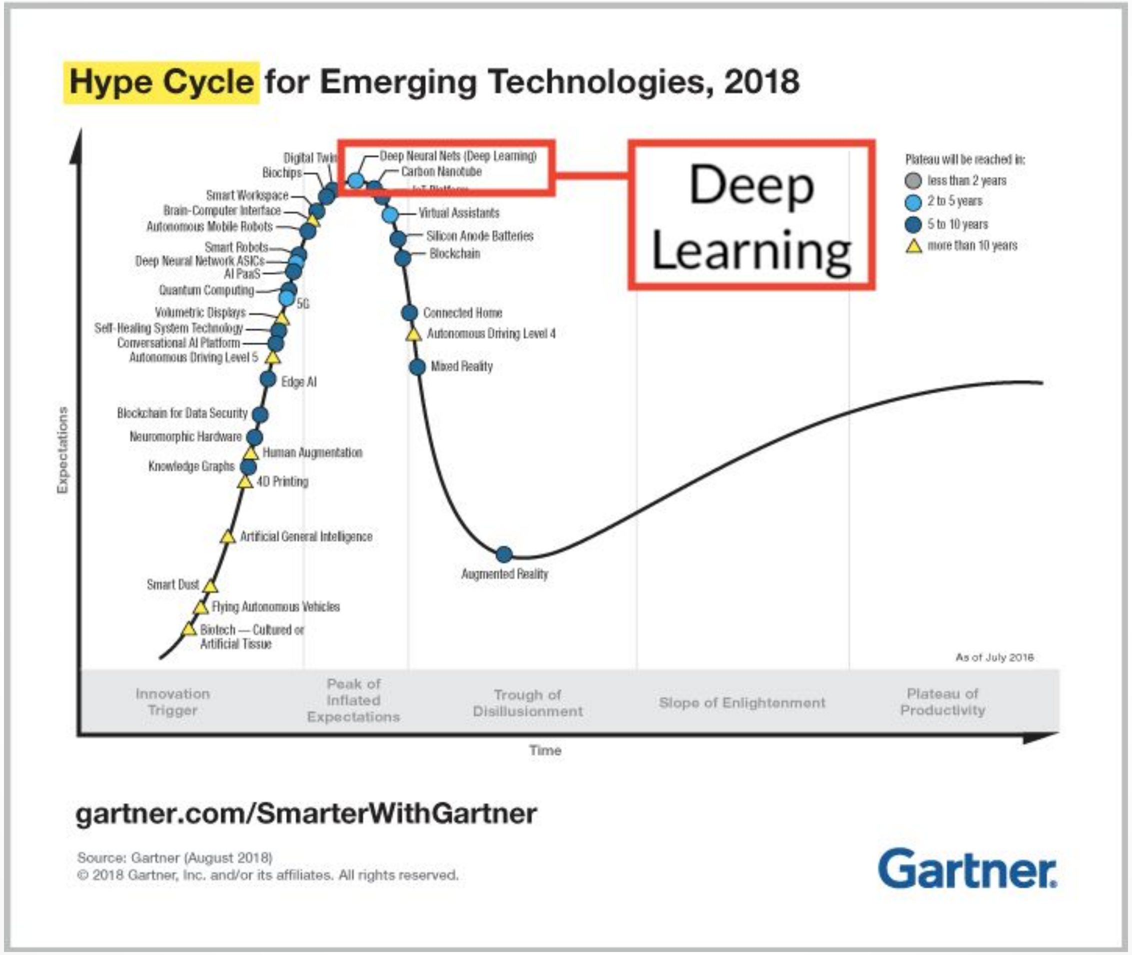
[Gatys et al. 2015]

Neural networks (GANs) as artists



This picture made by a GAN was sold for $432,500 and it's not even real.

Deep Learning and society expectation



Deep Learning's 'Permanent Peak' On *Gartner's Hype Cycle*

The Jim Cray's paradigms

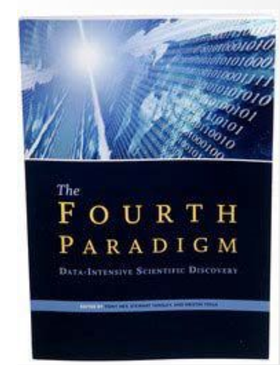
Science Paradigms

- Thousand years ago:
  science was **empirical**
  describing natural phenomena

- Last few hundred years:
  theoretical branch
  using models, generalizations

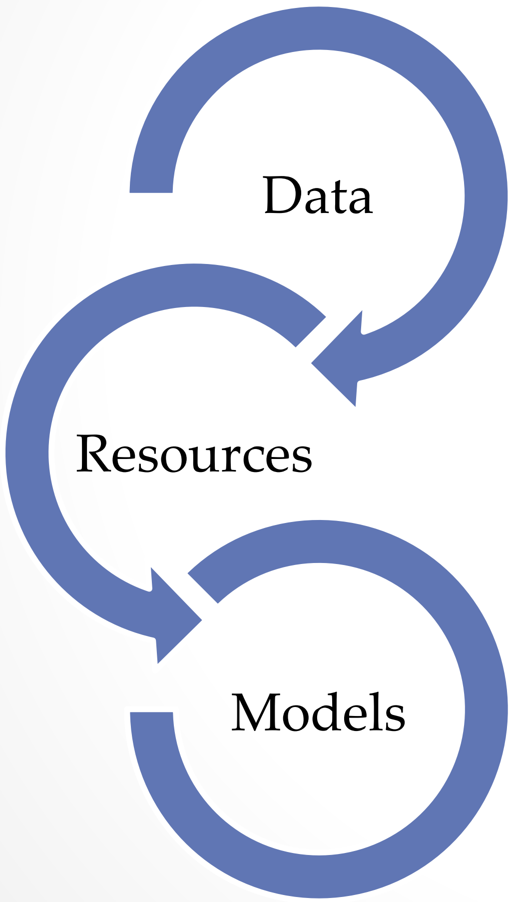
$$\left(\frac{\dot{a}}{a}\right)^2 = \frac{4\pi G\rho}{3} - K\frac{c^2}{a^2}$$

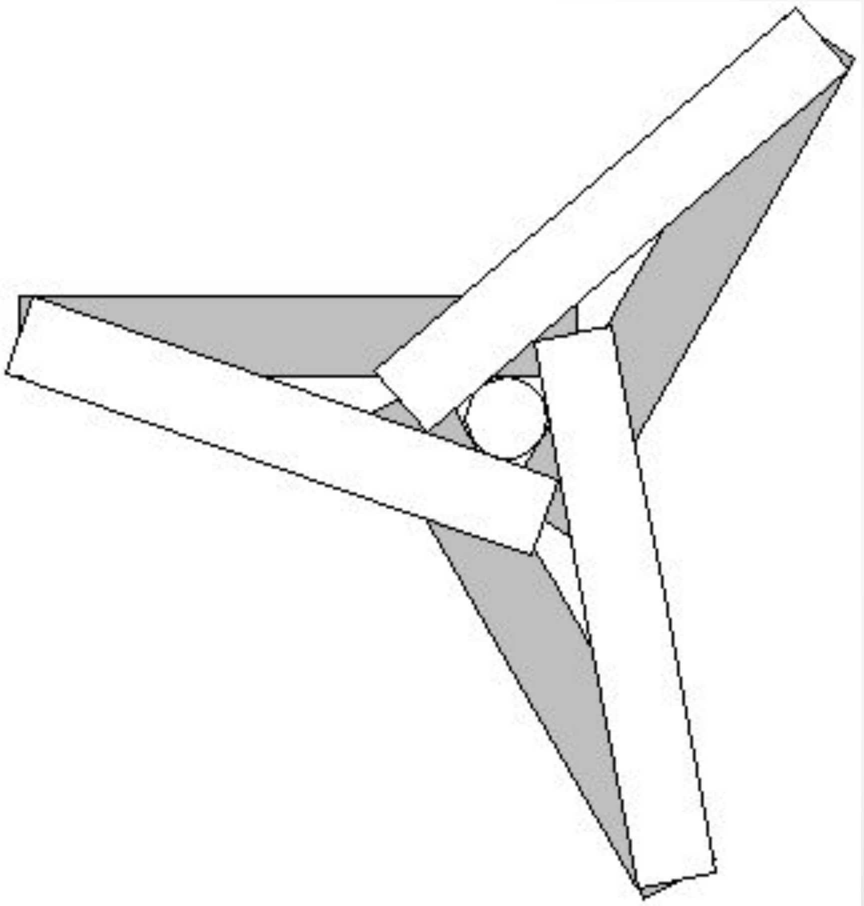
- Last few decades:
  a **computational** branch
  simulating complex phenomena
- Today: **data exploration** (eScience)
  unify theory, experiment, and simulation
  - Data captured by instruments
    or generated by simulator
  - Processed by software
  - Information/knowledge stored in computer
  - Scientist analyzes database/files
    using data management and statistics



Toni Hey, 2009

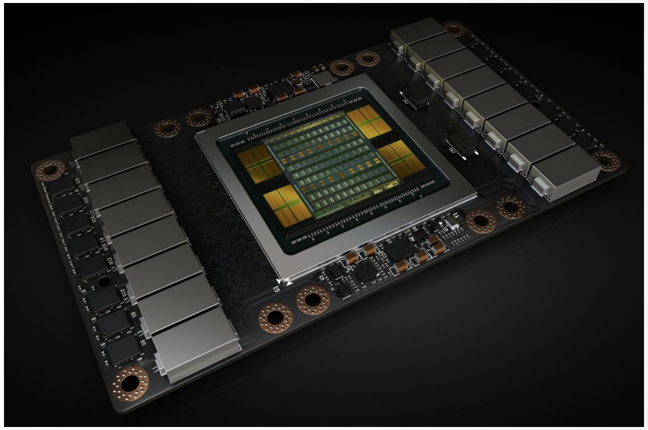
The magic triangle

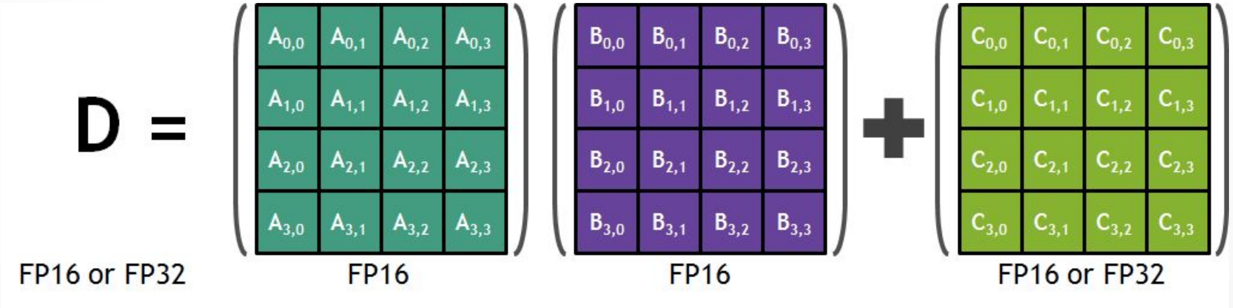




The Importance of GPUs

- Nvidia Tensor Cores - 2017
- Google Tensor Processing Unit (TPU) - 2016
- Intel - Nervana Neural Processor - 2017
- GPUs in Cloud Computing (Google, 2017)





GPU cores is based on matrix multiplication

https://www.doc.ic.ac.uk/~jce317/history-machine-learning.html#top

Data

90% of all digital data were generated last 2 years.

Every minute of the day:

- 4M YouTube videos watched
- 456K tweets on Twitter
- 46K potos posted in Instagram
- 16M text messages sent
- 103M spam emails sent

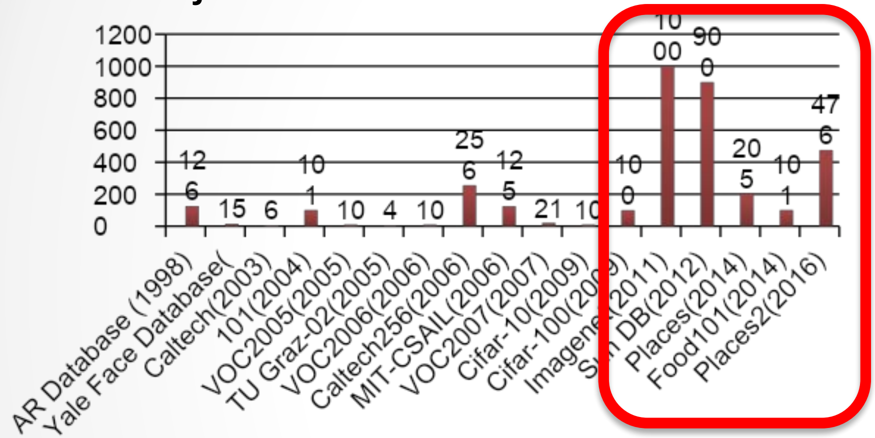
Daily:

- 300M photos get uploaded
- 95M photos and videos are shared on Instagram
- 100M people use the Instagram “stories”
- 15K GIFs are sent via Facebook
- 154K calls on Skype
- 4.7T photos stored in cameras

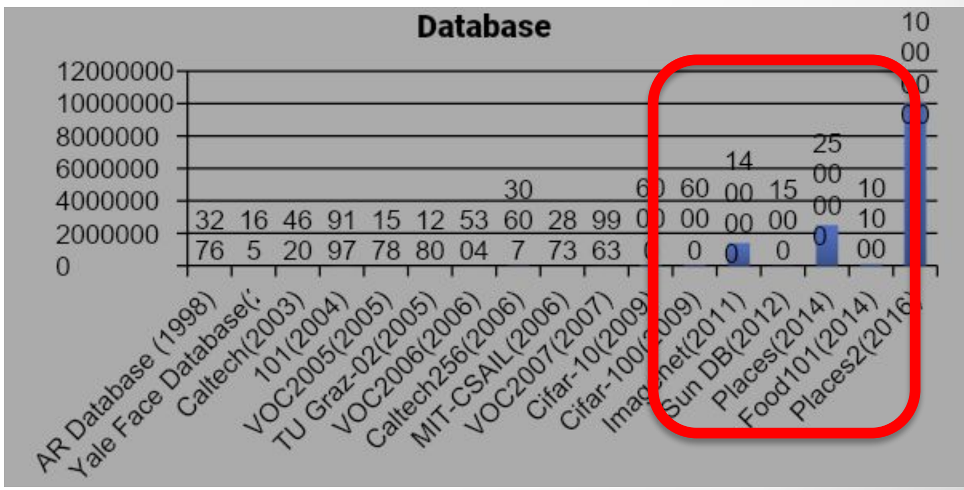
https://www.forbes.com/sites/bernardmarr/2018/05/21/how-much-data-do-we-create-every-day-the-mind-blowing-stats-everyone-should-read/#46be238160ba

Image databases evolution

Number of
objects/Database



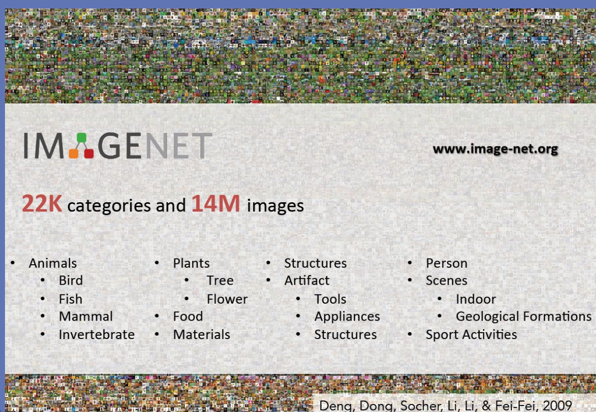
Number of images/Database



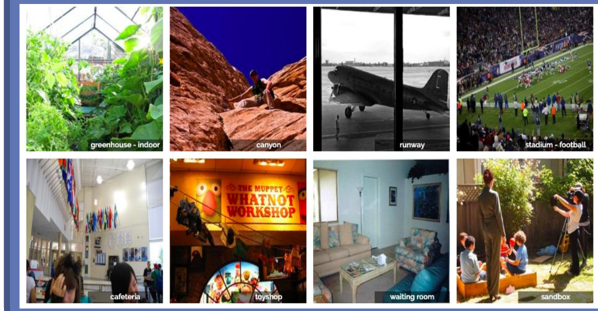
ImageNet & Deep learning

09:56

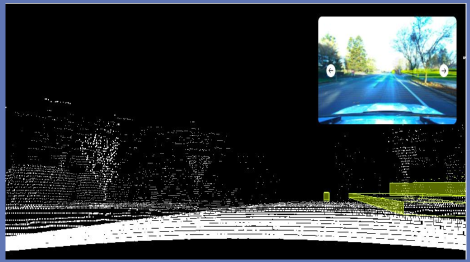
Deep Learning Datasets



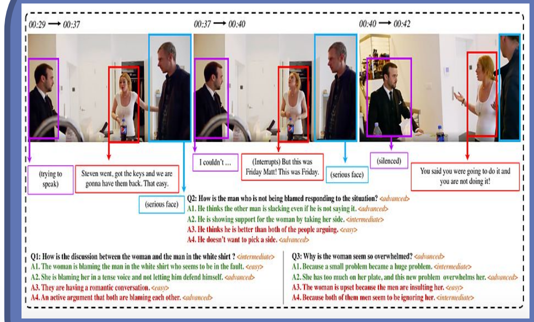
LVIS Challenge: 2.2M masks, 16K images



Places2: 10M images



[Lyft Level 5]



SocialIQ

TACO: Waste in the wild

FastMRI

Food datasets

Food256: 25.600 images (100 images/class)
Classes: 256

Food101 – 101.000 images (1000 images/class)
Classes: 101

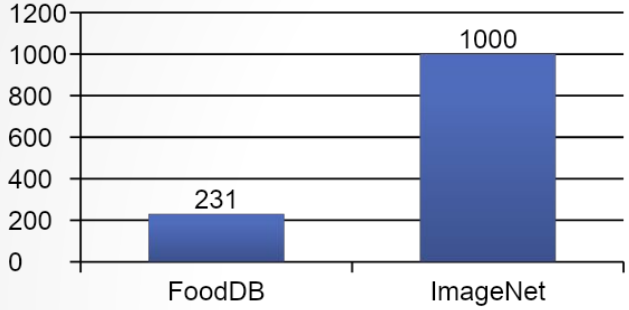
Food101+FoodCAT: 146.392 (101.000+45.392)
Classes: 231

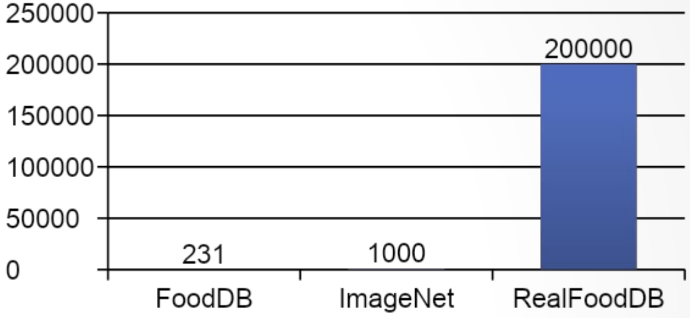
Food DB	ImageNet	Future Food DB
150.000 images 231 categories	1.400.000 images 1000 categories	????? images 200.000 categories

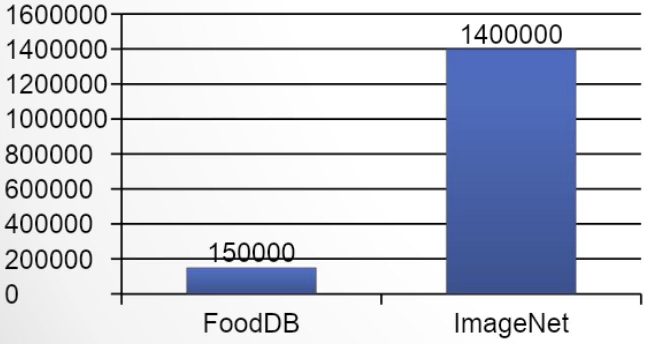
FoodImageNet soon to come!

09:56

How many images should contain the real FoodDB?



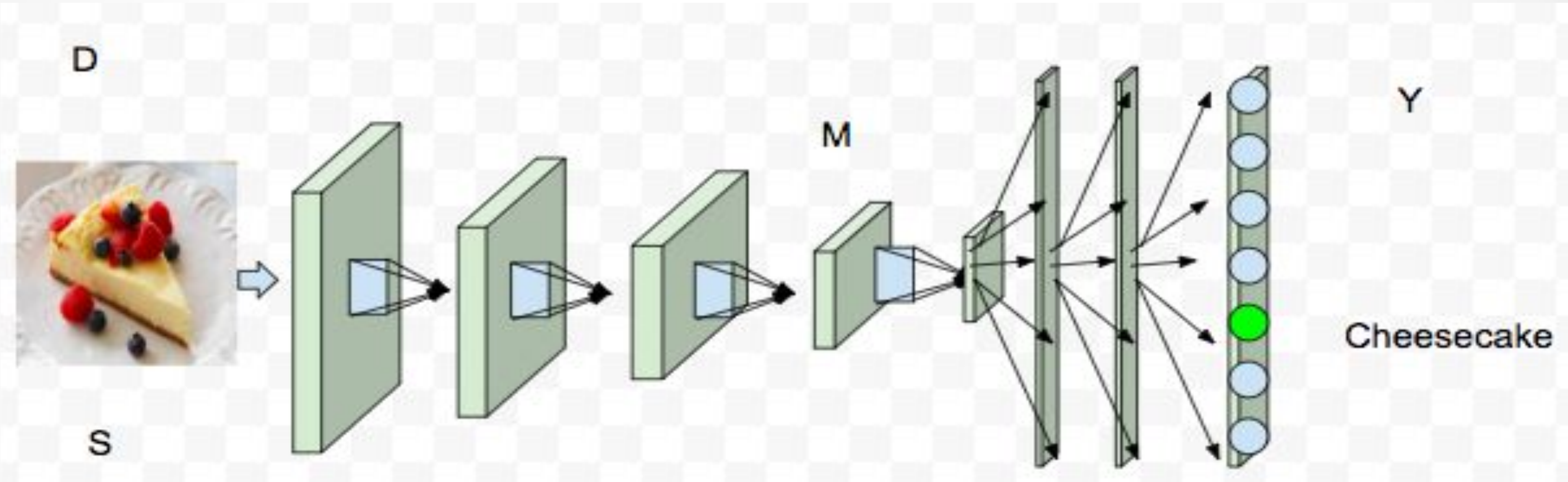


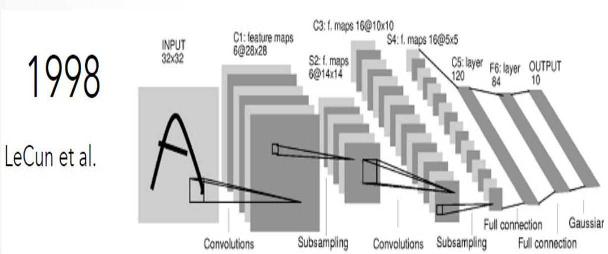


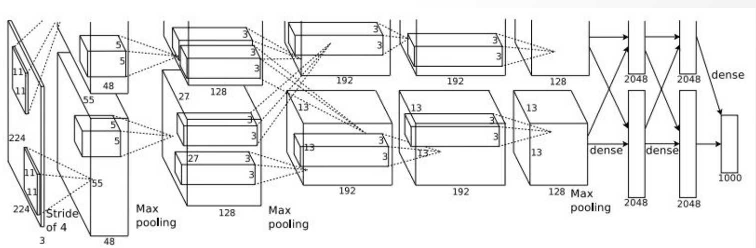
09:56

One is for sure,
if there is a solution,
it is highly probable to need
Deep learning!

What is a Neural Network?



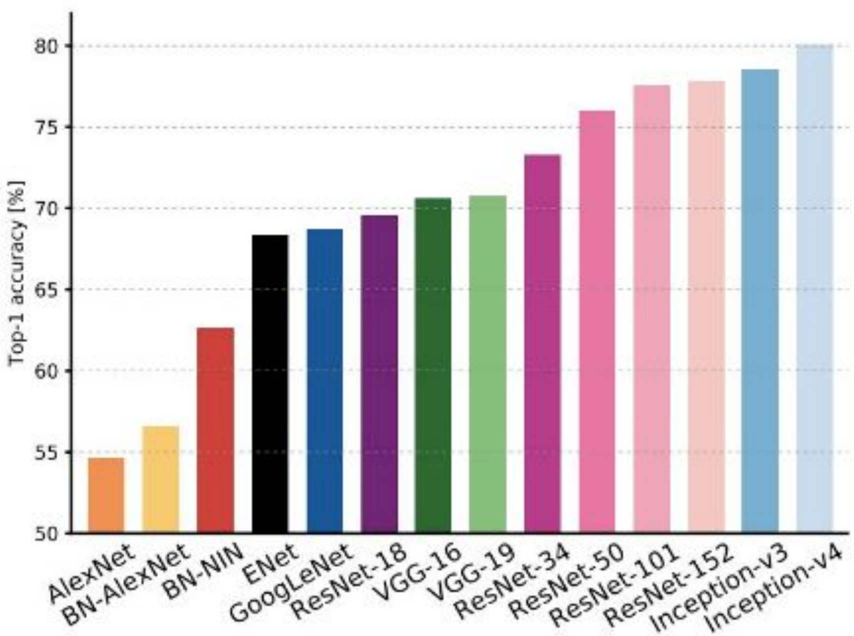


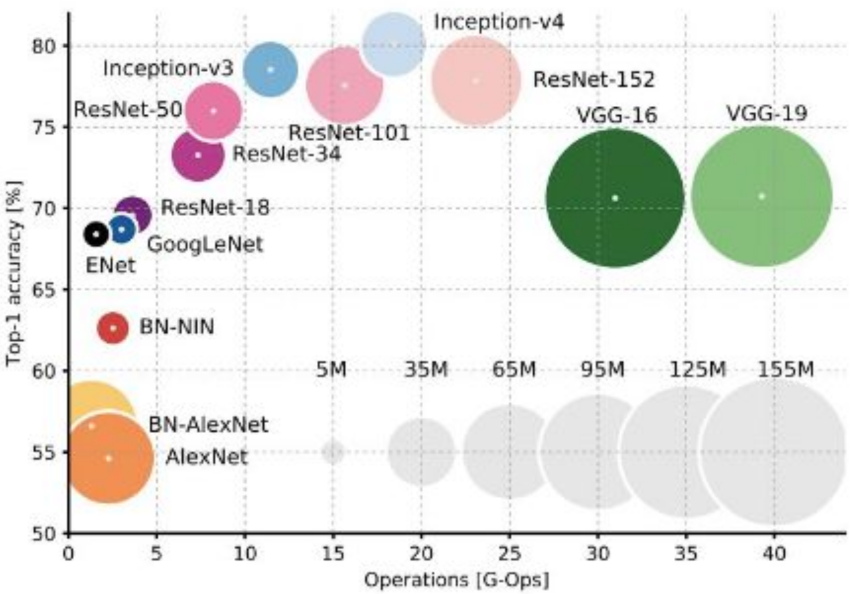


LeCun, Chief AI Scientist for **Facebook** AI Research (FAIR), and a Silver Professor at New York University

A.Krijevksi et.al. 2012, Google Brain & Waymo.

Analysis of CNNs





- **Millions of parameters!!!**

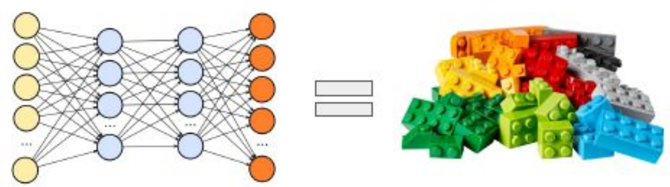
The process of training a CNN consists of training all hyperparameters: convolutional matrices and weights of the fully connected layers.

What makes DNN so popular?

It has the three advantages:

- 1. Self-learned high-level features representations

- 2. Modularity



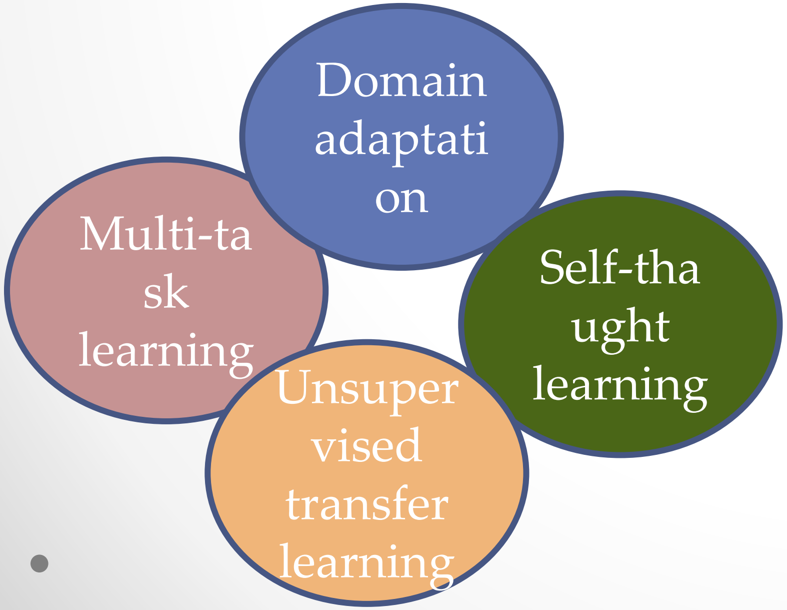
- 3. Transfer Learning

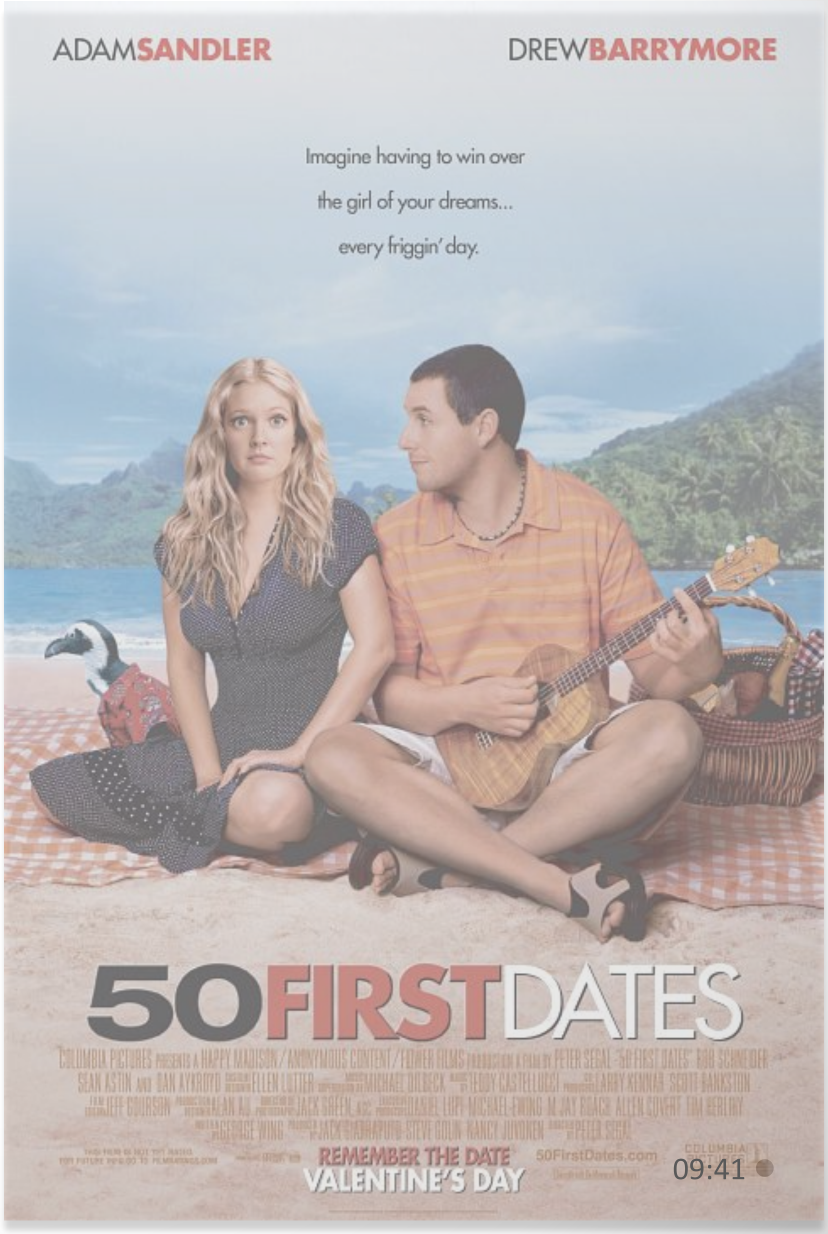
Use Transfer Learning

Henry Roth is a man afraid of commitment up until he meets the beautiful Lucy.

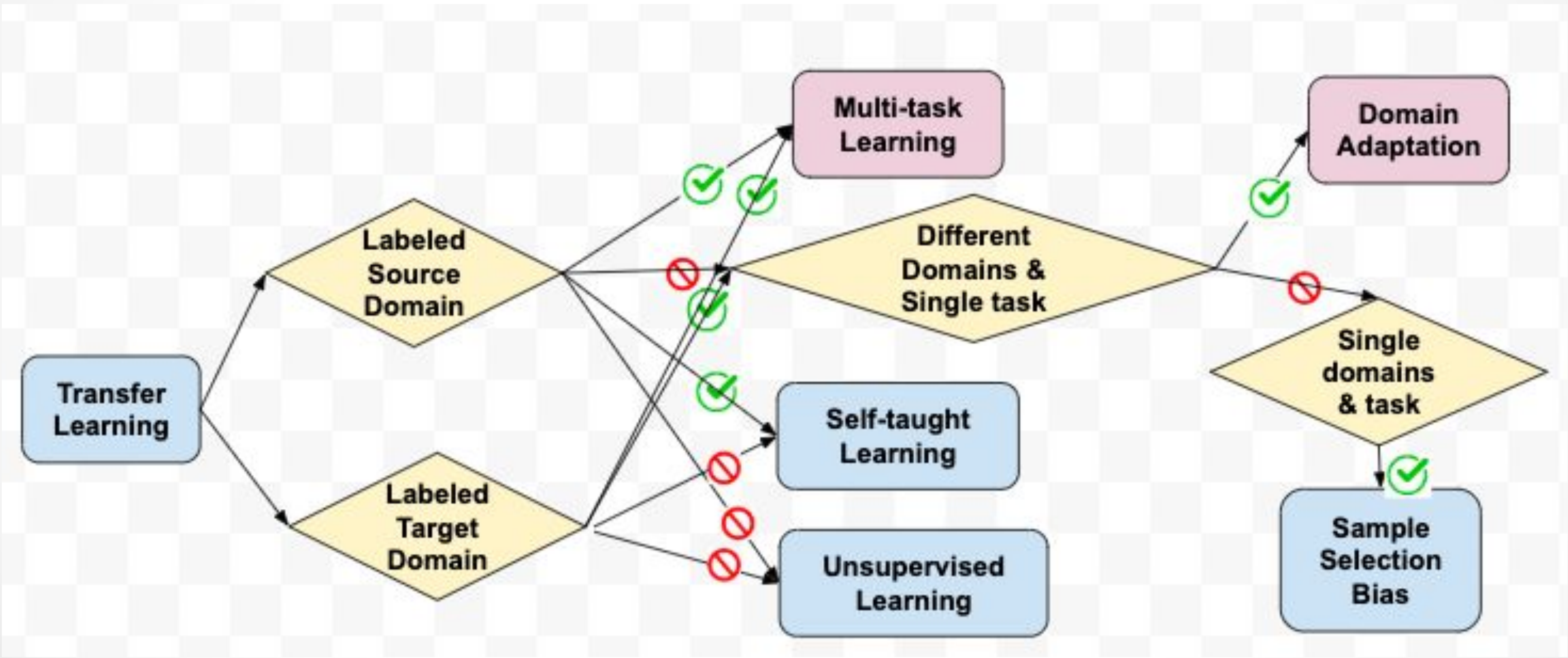
They hit it off and Henry think he's finally found the girl of his dreams,

until he discovers **she has short-term memory loss and forgets him the next day.**

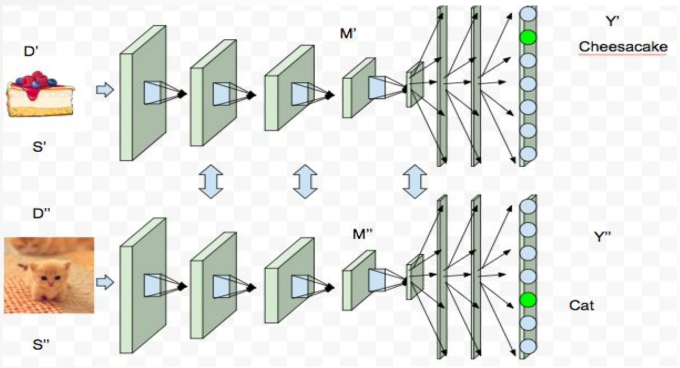


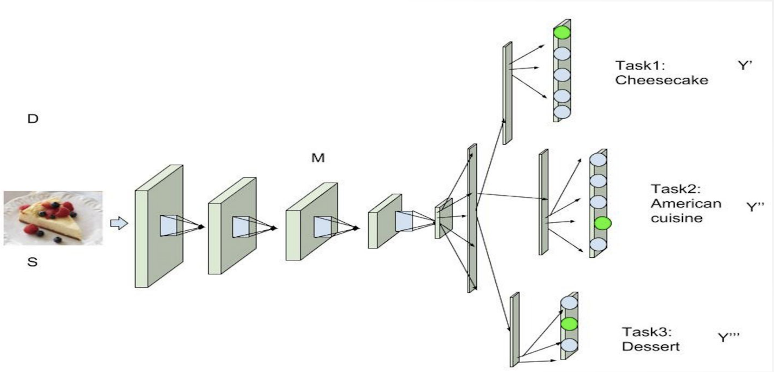


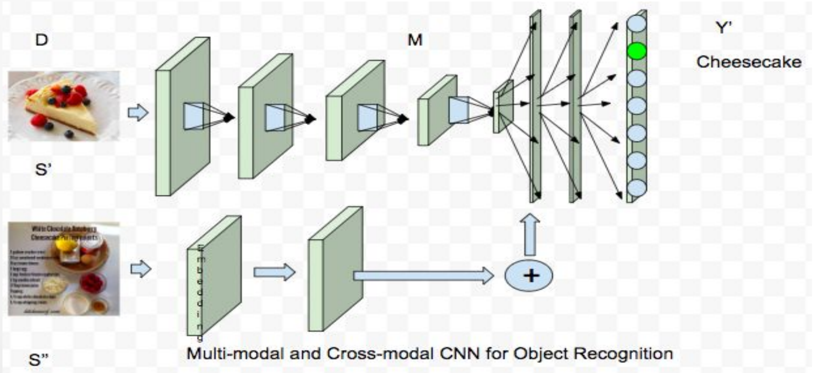
Transfer Learning

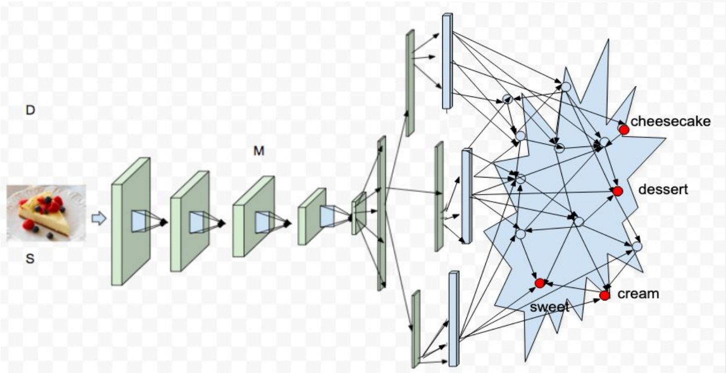


Transfer learning (TL)









Food Recognition as 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.

Multi-Task Learning (MTL)

- Learning **multiple objectives** from a shared representation

  - *Efficiency* and prediction *accuracy*.

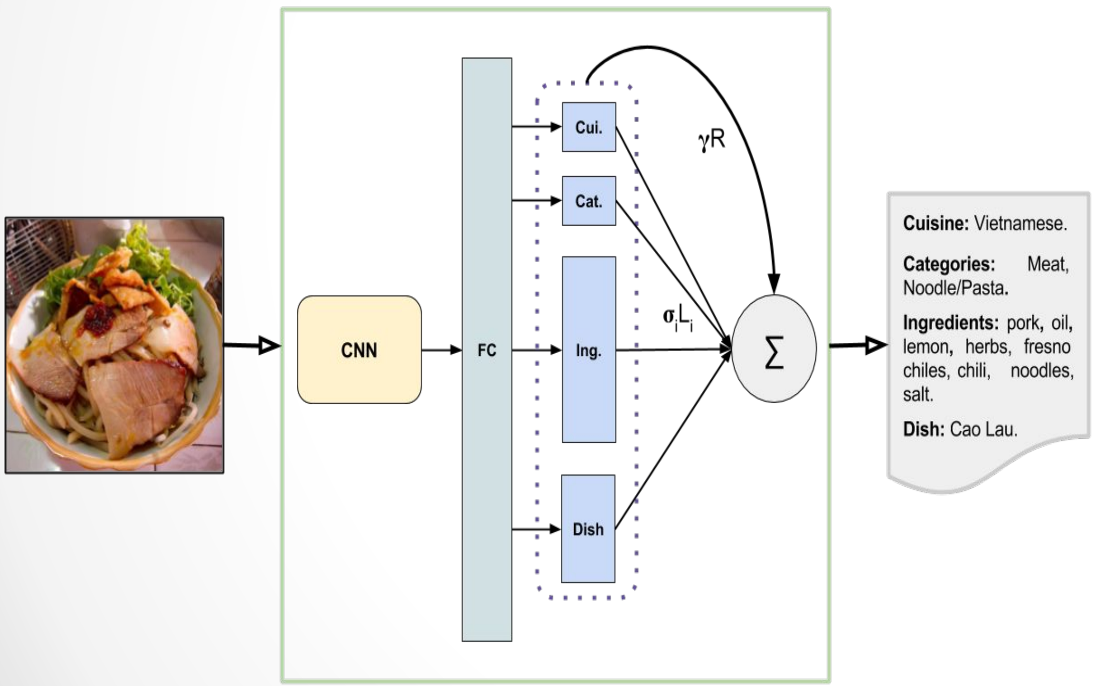
- Crucial importance in systems where **long computation** run-time is prohibitive

  - Combining all tasks *reduces* *computation*.

- Inductive **knowledge transfer**

  - *Generalization* 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

But many unanswered questions...

- Why doesn't my model work?
- -> Why does my model work?
    - Why does my model work?
    - What does my model know?
    - Why does my model predict this and not that?

- **Our models are black boxes and not interpretable...**
    - Physicians and others need to understand why a model predicts an output.

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)

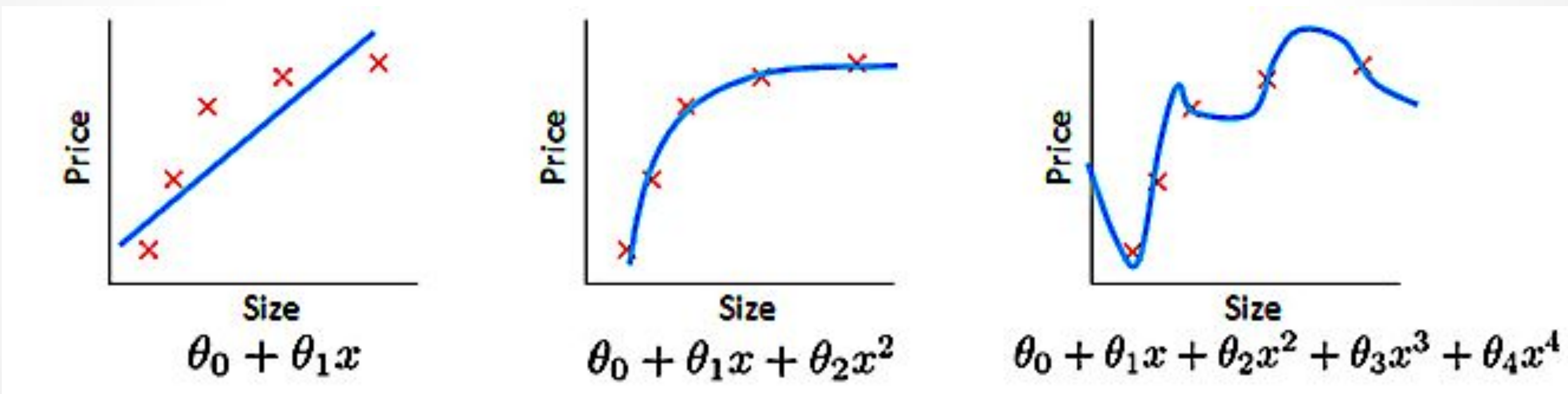
Model uncertainty

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?



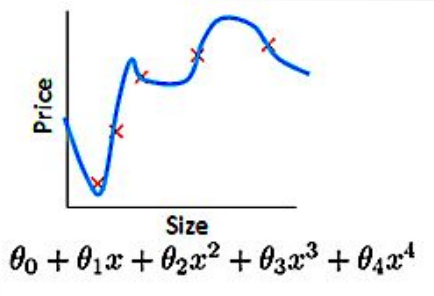
Gal (2016)

Types of uncertainty in Bayesian modeling

Aleatoric – captures the noise inherent in the observations

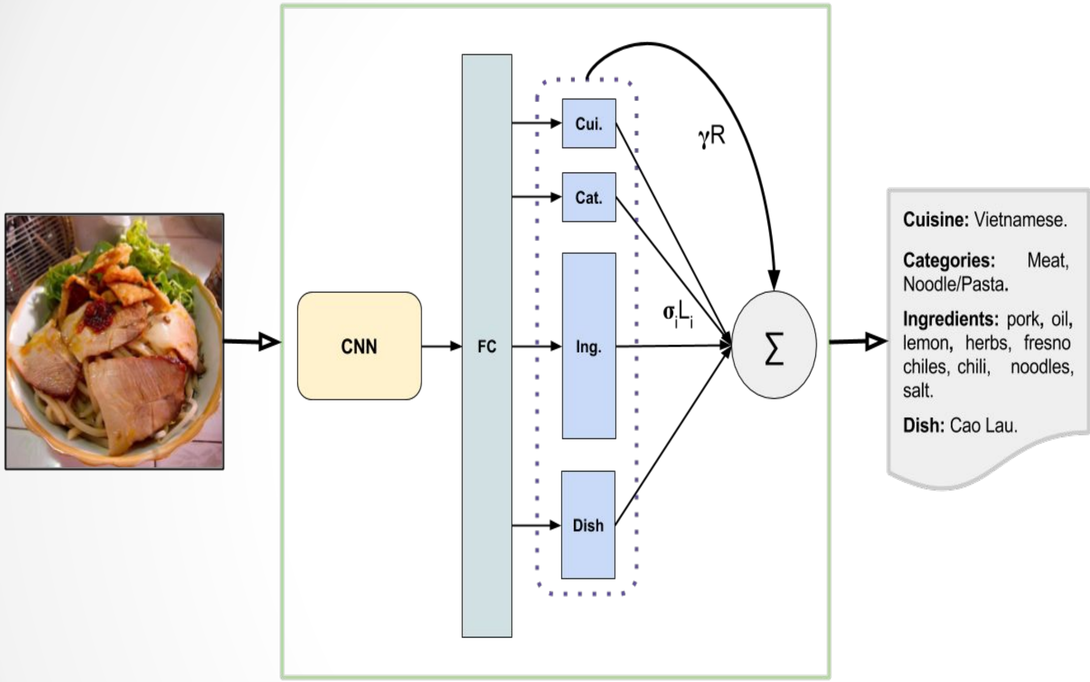
- heteroscedastic – data-dependent
- homoscedastic – constant for different data points,
  - but can be task-dependent.

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



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 more clever!

Our FoodImageNet

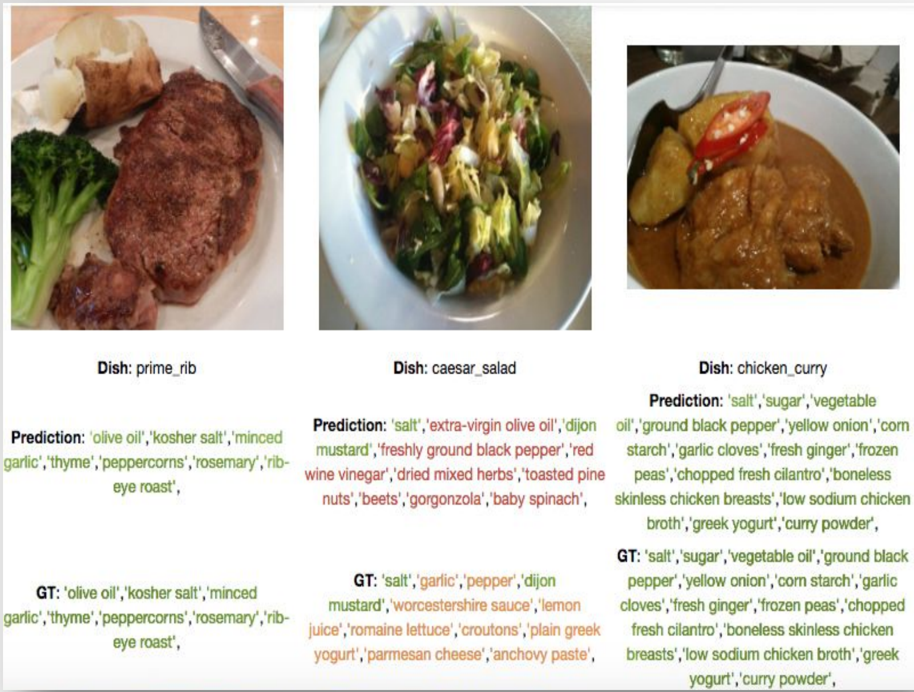
Our FoodImageNet

- Food – 450 dishes, 11 categories, 11 cuisines
- Ingredients – 65
- Drinks – 40
- Labeled images
- Segmented images
- Recipes

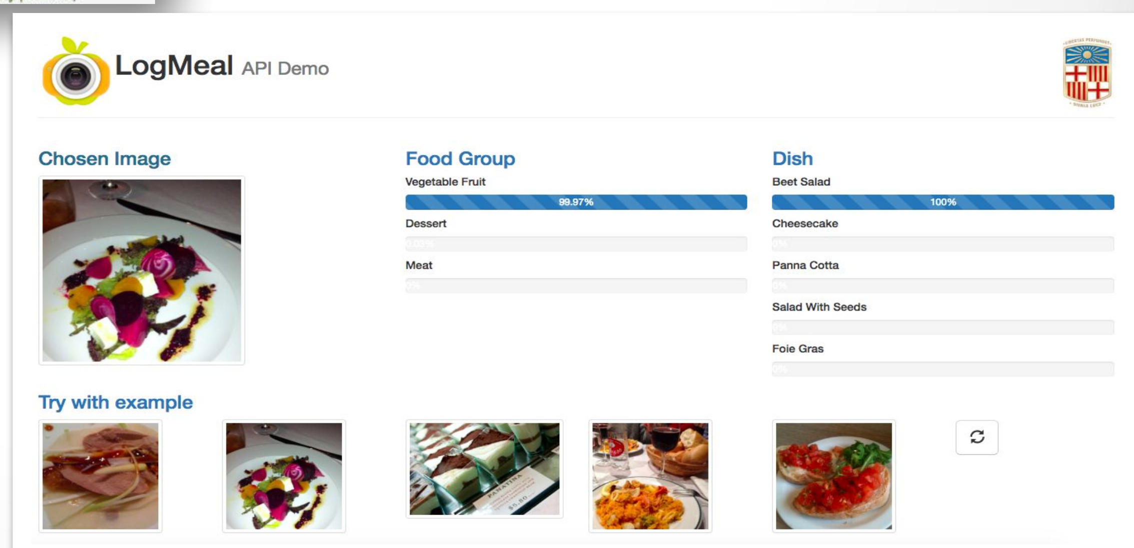
In total:
more than
550.000 images

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)

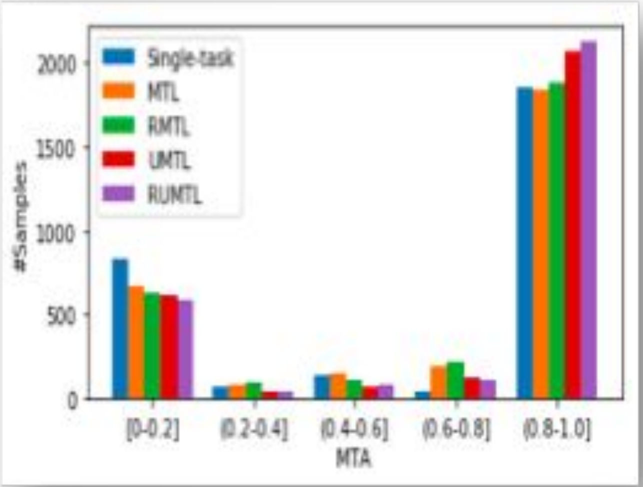
Food ingredients recognition

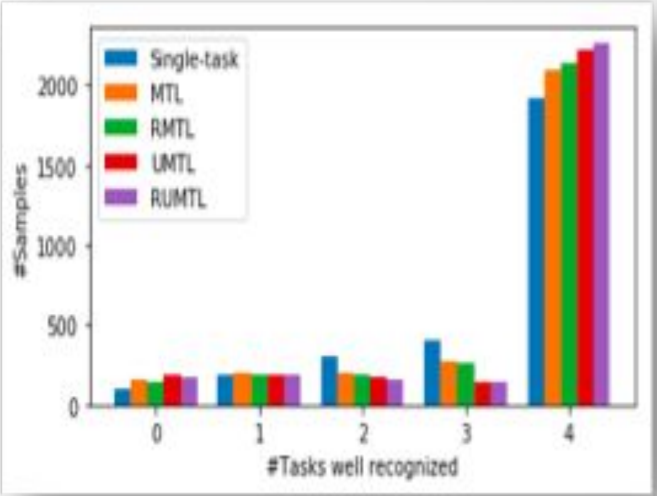


Food category and class recognition



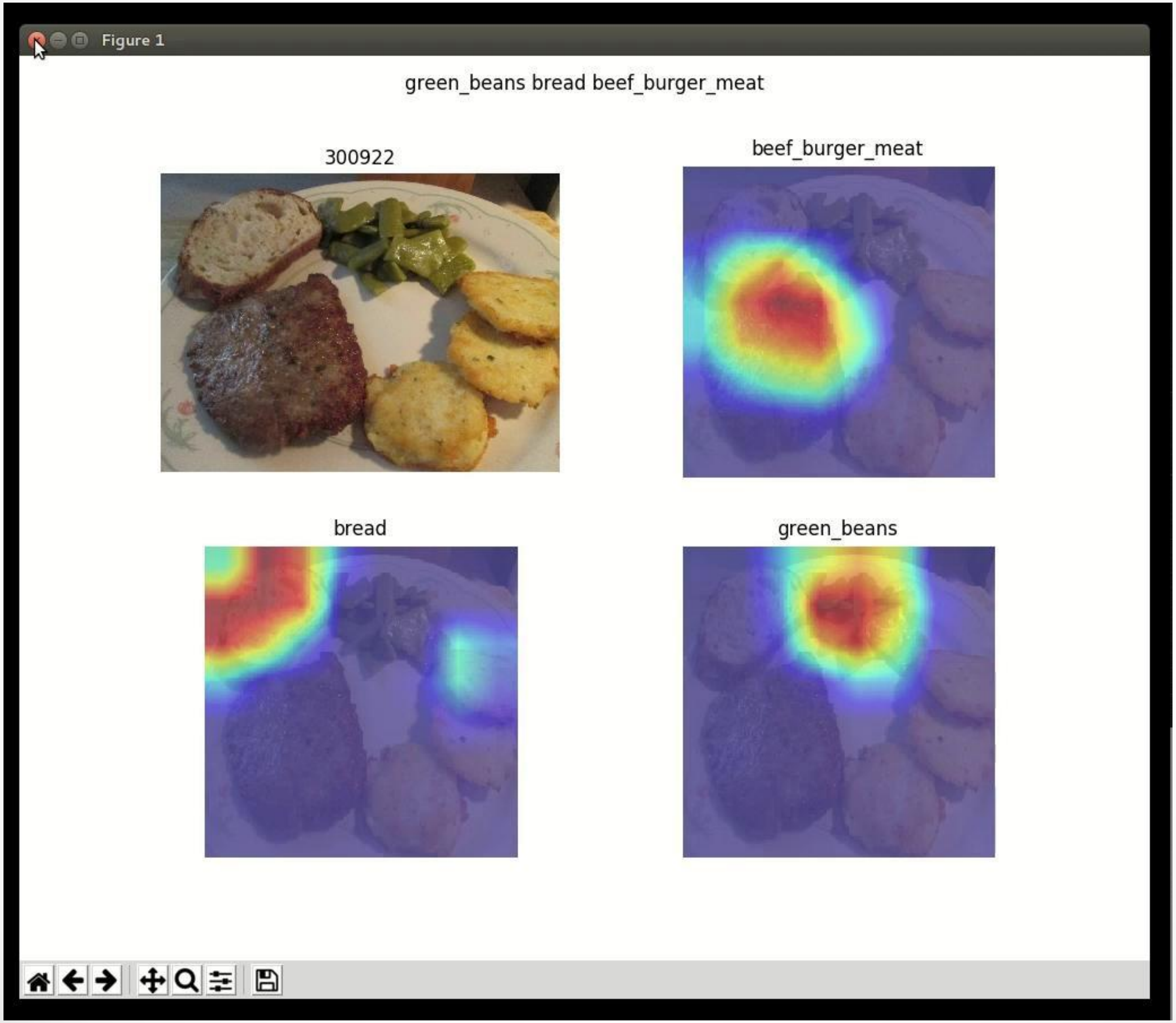
Food Recognition





	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**

Food Recognition



Understanding the cooking process

By Mostafa Kamal, Domenec Puig et.al.

Conclusions

- Food image world brings us huge amount of data and Computer Vision questions
- It makes us redefine which are :
  - Datasets
  - Problems, Q&A
  - Methodologies & Technologies
- Transfer learning and its subproblems as multi-task learning open huge amount of opportunities
- Uncertainty modeling is a hot topic with many open questions and challenges!
  - Epistemic uncertainty
  - Aleatoric uncertainty
- A huge impact of food analysis is expected from point of view of:
  - Science, but also
  - Real world applications, specially important for the society.


Thank you!