The Global Language of Business

Belgium & Luxembourg

GS1 Innovation Café

Artificial Intelligence for better data quality

14 November 2019

Agenda

15:00 - Welcome by the chairman of the GS1 Innovation Committee
Wouter Schaekers - Supply Chain Innovation & Sustainability Procter & Gamble

15:05 - AI and Data Quality
Jan Merckx - Innovation Manager, GS1 Belgilux
Mayra Castellanos - Innovation Manager, GS1 Germany

15:30 - From the GS1 Research Labs
Klaus Fuchs - Auto-ID Labs, ETH Zürich

16:00 - The use of AI at GS1
Eelke van der Horst - Data Scientist, GS1 Netherlands

16:15 - Smart with Food
Ellen Verhasselt - International Business Development, Smart With Food

16:30 - Q&A
17:00 - Networking reception

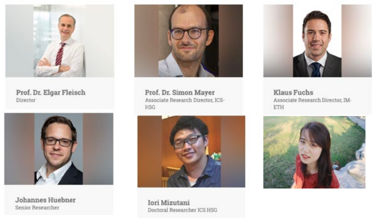
Artificial Intelligence for better data quality.
Can AI improve data quality?

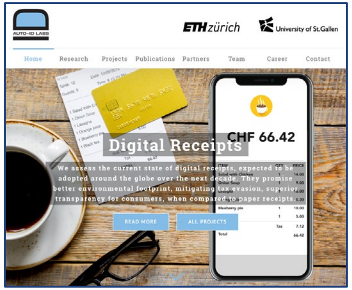
Klaus Fuchs
Auto-ID Labs ETH/HSG, ETH Zurich

Auto-ID Labs ETH / HSG: Our Chair combines research on internet of things in the retail & health domain (consumer health)

Auto-ID Labs ETH/HSG (Focus: Retail)

- Focus on research on IOT in the retail domain. Strong partnership with GS1 - the global standards organisation.
- Team: 2 Professors, 1 PostDoc, 4 PhD candidates (Lead: Klaus Fuchs)
- Auto- ID Labs as a network of research centers around the globe: MIT, ETH, Keio, KAIST, Tongji, Cambridge
- www.autoidlabs.ch
- www.autoidlabs.org





Center for Digital Health Interventions (Focus: Health)

- Focus on digital health interventions, incl. mobile coaches, chatbots and automatic sensing of health statuses in asthma, diabetes. Partnership with CSS - the largest Swiss health insurance.
- Team: 4 Professors, 9 PhD candidates (Lead: Dr. Tobias Kowatsch)
- www.c4dhi.org/

Agenda for the AI Innovation Café on Nov 14th, 2019

1. Identifying product category from product images
2. Text mining to infer product categories for Nutri-Score
3. Identifying products within retail environments via computer vision

- **Motivation:** Academic Research & GS1 Partnership

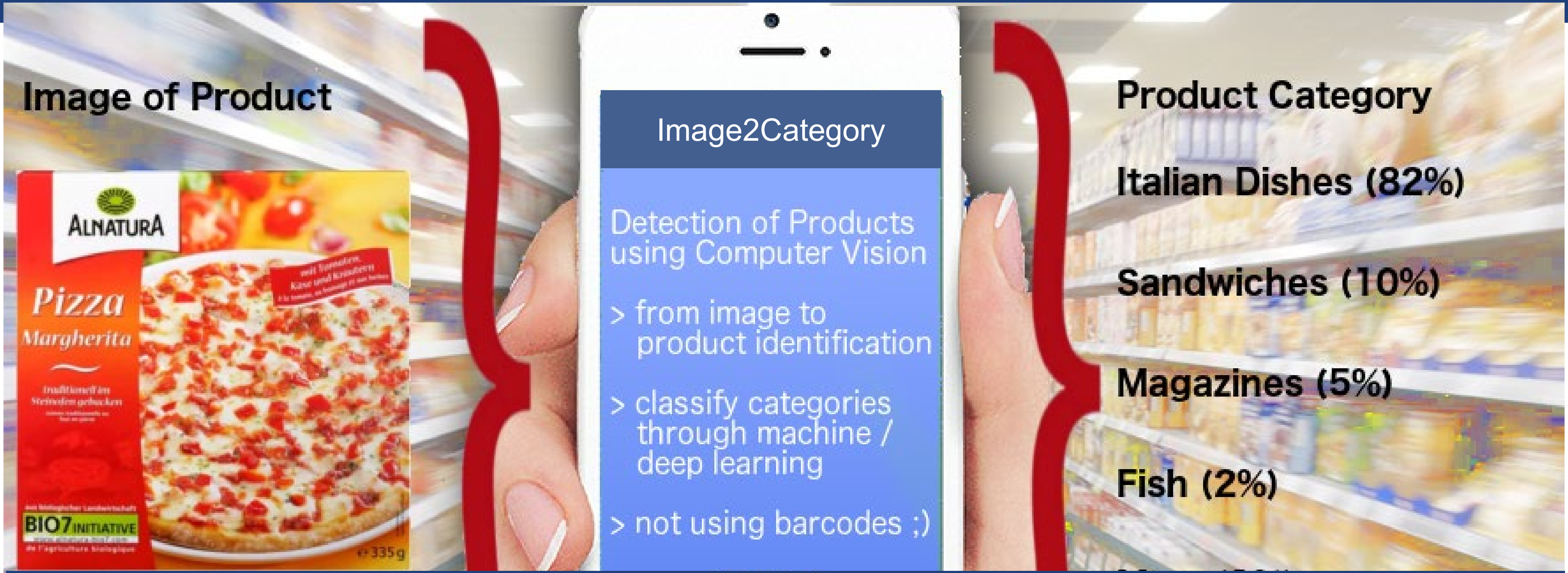


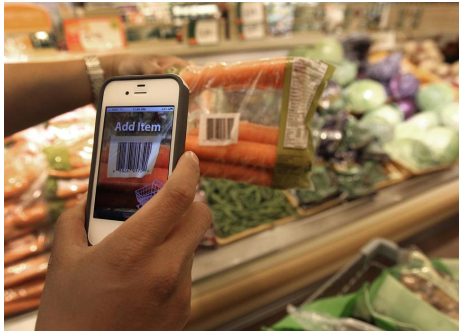
Image2Category

Can we infer a product category from a single product image? A typical data quality task.

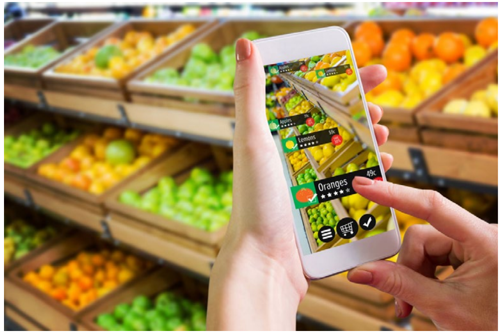
Motivation: Identifying product category from product images

- Support transparency and customer buying experience. For example, labels and warnings, orientation in the supermarket, ecommerce categorization

Now: Barcode Scanning



End Goal: Shelf Scanning



Future: AR/VR -

Image2Category: Data Preparation

- Dataset collected from GS1’s database Trustbox
- Initial dataset: 15’891 product images
- 19 major categories
- 106 (minor) sub-categories
- Out of 15’981 only 3’211 labeled images
  > only major categories
- Split train/test data: 80/20 %

Example of product images

Image2Category: Data Preparation

- Manual (!) labelling by us, so we started with 3k images.

ID	Category	Number of Images
1	Soft drinks	322
2	Alcoholic drinks	0
3	Breads, flakes and breakfast cereals	159
4	Eggs	0
5	Fats and oils	85
6	Fish	97
7	Meat substitutes	24
8	Meat and offal	137
9	Processed meat and sausages	136
10	Fruit	83
11	Vegetable	194
12	Courts	361
13	Cereal products, legumes and potatoes	165
14	Milk and milk products	255
15	Nuts, seeds and oil fruits	21
16	Salty snacks	87
17	Special food	3
18	Sweets	686
19	Miscellaneous	396
	Total	3211

ID	Category	Number of Images
1	Soft drinks	322
2	Breads, flakes and breakfast cereals	159
3	Fats and oils	85
4	Fish	97
5	Meat, offal, processed meat and sausages	273
6	Fruit	83
7	Vegetable	194
8	Courts	361
9	Cereal products, legumes and potatoes	165
10	Milk and milk products	255
11	Salty snacks	87
12	Sweets	686
	Total	2767

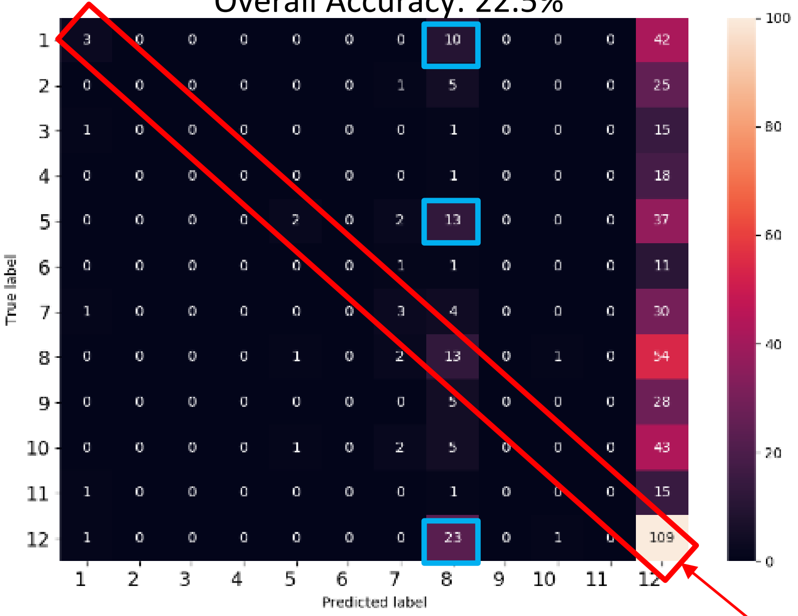
→ 80/20 % train/test split

→ 2214 training images/553 test images

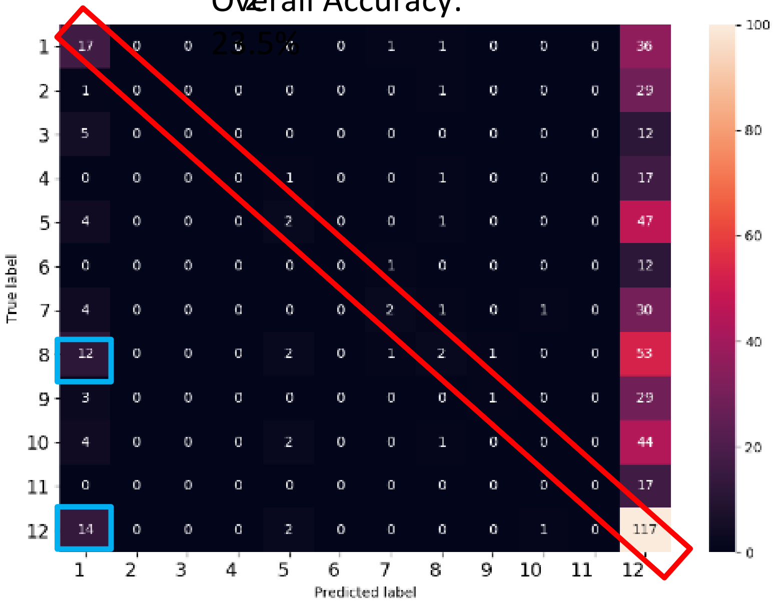
Image2Category: Why does Deep Learning (CNN) fail

- Counter-intuitive, why doesn't it work?

MobileNetV2
Overall Accuracy: 22.5%



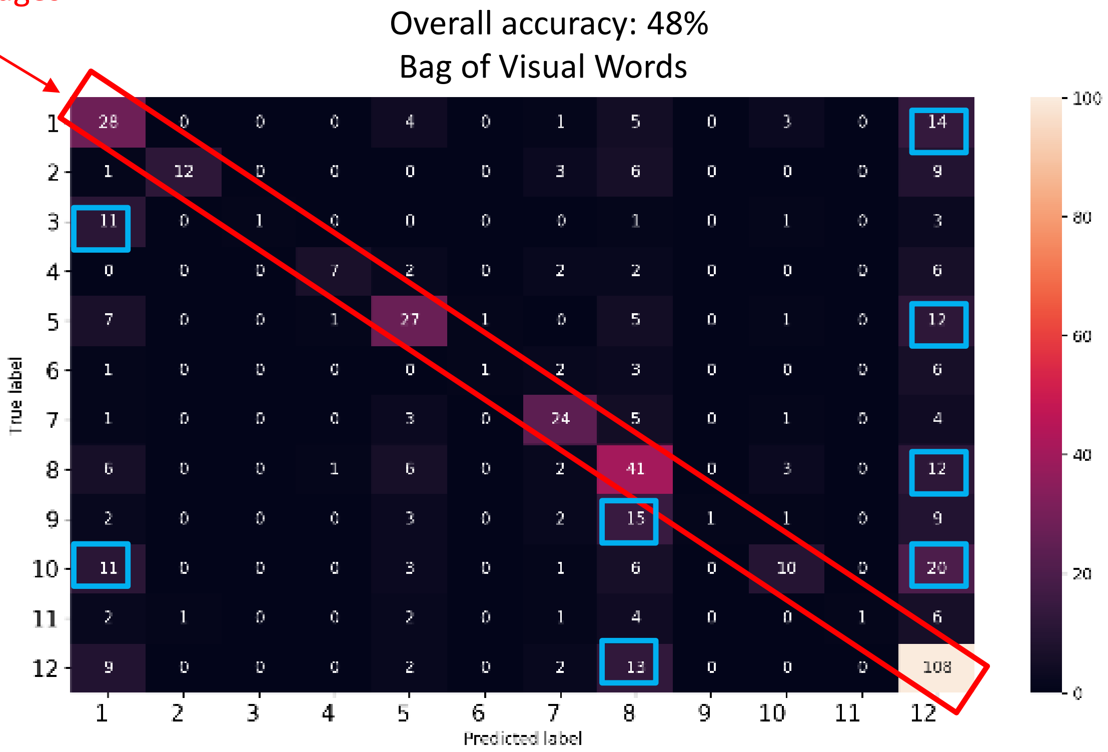
InceptionResNetV2
Overall Accuracy: 23.5%



correctly classified images

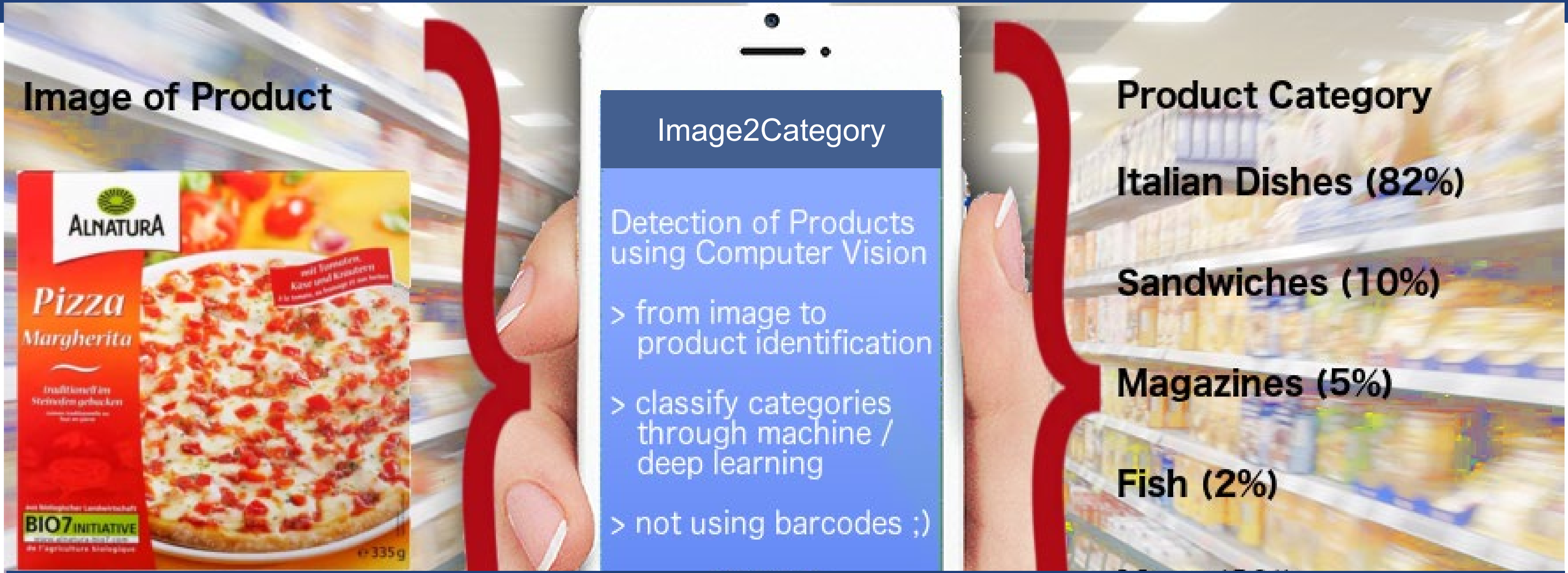
Image2Category: Bag of Visual Words better

correctly classified images



Per Class Accuracy:

(1) Soft Drinks: 50.9 %
(5) Meat: 50.0 %
(7) Vegetables: 63.2 %
(8) Courts: 57.7 %
(12) Sweets: 80.6 %



Why did Image2Category fail?

Learning from failure

Image2Category: Why did we fail?

- Overall accuracy = correctly classified images/total # images (553 test images)
- CNN models need a lot of data to perform well → poor performance of CNN models due to a low number of image data
- “Sweets contain a great variety of packaged goods” → hard to distinguish → Bias towards “Sweets”
- Bag of Visual Word can cope better with fewer data.
- Less biased towards “Sweets”
- A lot of room for improvement…

Image2Category: Why did we fail?

- Inconsistent/arbitrary classes, and overlap of products: is a marinated steak also processed meat or not?

Processed meat or not?

Processed meat

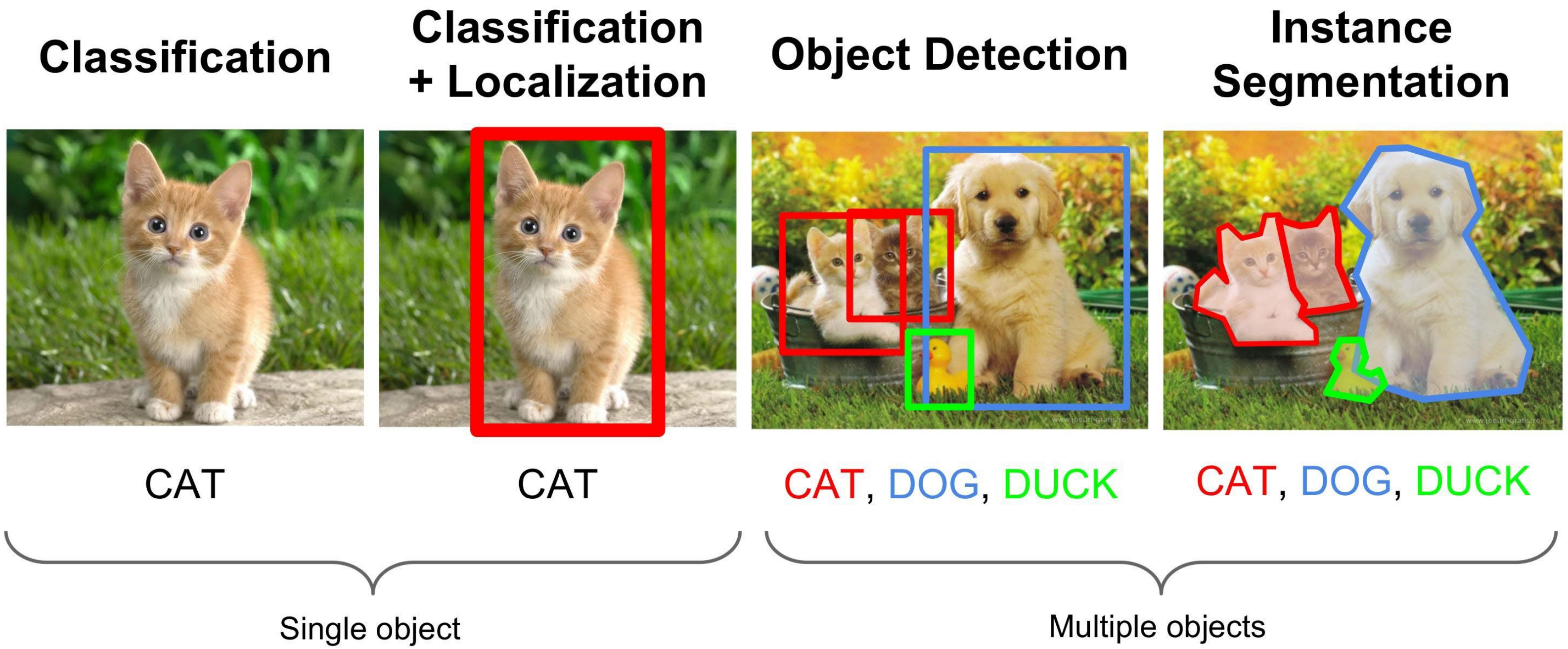
Image2Category: Why did we fail?

- Feature variety: Also, ecommerce/transparent images vs retail images, vs logos…

Example Images of category “Sweets”

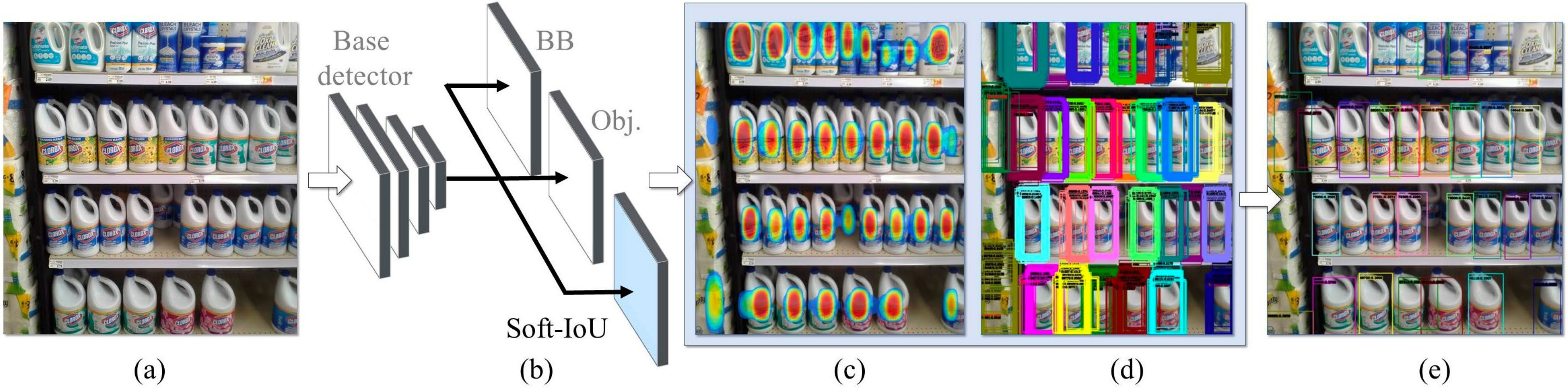
- Sparse data (at least in some classes)
  - Aim for >1000 per class
- Imbalanced dataset

Problem: Computer vision-based identification of packaged products is challenging (density, # of classes, feature variety)



[1] Google, 2019.

Problem: Computer vision-based identification of packaged products is challenging (density, # of classes, feature variety, occlusion, reflection, deformation)



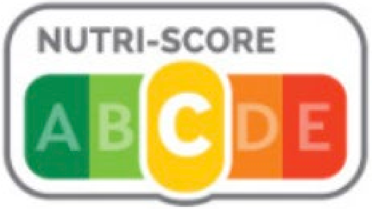
[1] Eran Goldman*, Roei Herzig*, Aviv Eisenschtat*, Jacob Goldberger, Tal Hassner, Precise Detection in Densely Packed Scenes, 2019.

Text2Category: Text mining to infer product categories

Can we calculate a correct Nutri-Score for a product for which we do not know the product category?

Motivation

- In order to calculate the Nutri-Score, we need to know if a product is… Mineral Water, Beverage, Cheese, Added Fat, Food, or No Such Item



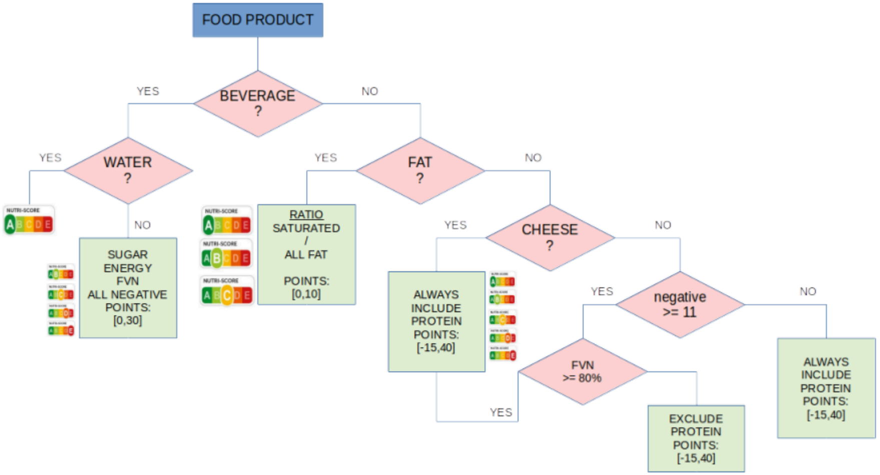


Figure 1.2: The flow char show the category dependent decisions that need to be done during the *Nutri*-score calculation. Generally the negative points are collected based on energy, sugars, fats and salt. Then the positive points from the FVN %, proteins and fibres. To conclude with the final score by subtraction of the positive from the negative.

Text2Category: Dataset

- Now, we had labelled 14k products
- We took ingredient texts & nutrients

Category	Number of products	Percentage
Beverages (0)	1119	7.9
Cheeses (1)	585	4.1
Fats (2)	415	2.9
Food (3)	11898	84.1
Water (4)	131	1.0

Table 2.2: Labelled product samples with respect to the categories.

Feature	Beverages (0)	Cheeses (1)	Fats (2)	Food (3)	Water (4)
energyKcal	942	585	355	11503	131
energyKJ	942	585	355	11503	131
totalFat	938	585	353	11368	131
saturatedFat	933	582	349	11206	131
totalCarbohydrate	938	555	351	11298	131
sugars	915	555	343	11247	131
salt	940	585	334	11257	131
sodium	940	585	334	11257	131
protein	937	585	353	11341	131
dietaryFiber	895	549	269	10611	128
ingredient_de	850	283	303	8682	78
ingredient_en	181	60	62	1911	8
ingredient_fr	502	166	185	5037	30
ingredient_it	418	125	142	4265	3

Table 2.3: Overview full data set that have a label with respect to the features of interest (14148)

Text2Category: Frameworks

- TF-IDF [Ingredient Texts]
- TF-IDF Weighted (Linear) [Ingredient Texts]
- **TF-IDF Weighted (Exponential) [Ingredient Texts]**

- Random Forest [Nutrients]
- **Stacked Random Forest [Nutrients]**

- **Combination of TF-IDF Weighted (Exponential) [Ingredient Text] & Stacked Random Forest [Nutrients]**

Text2Category: Minimum Confidence

- **If we include a minimum confidence for a classification, only few products are 'lost' but accuracy is better**

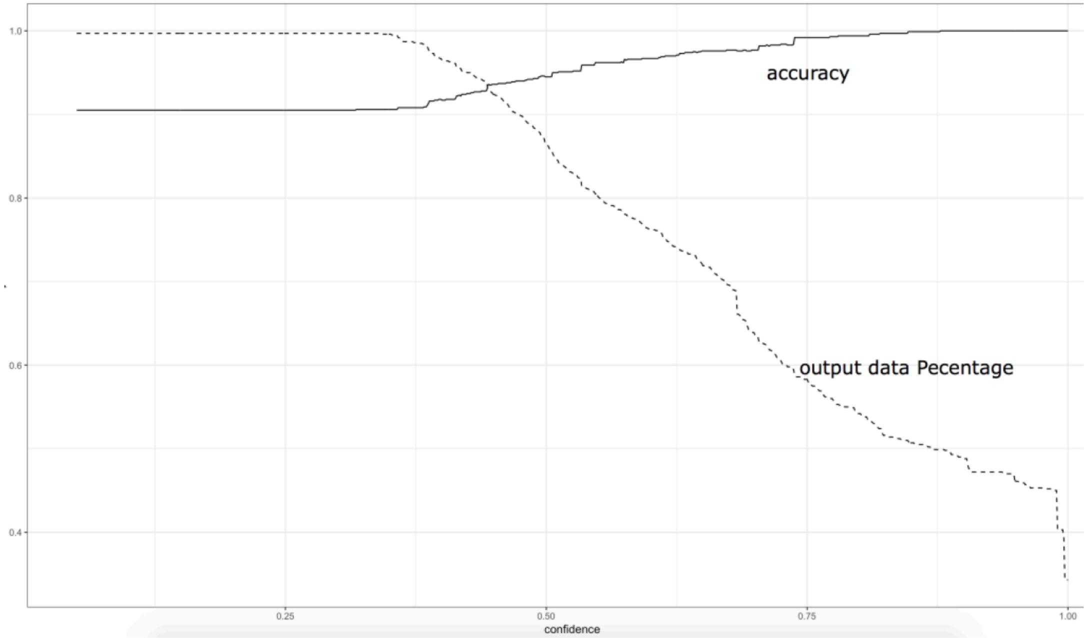


Figure 6 Trade-off between accuracy rate and inclusion rate as response to minimum confidence threshold.

Text2Category: Results

- Very successful results, across languages and with different models.
- **Very promising that Nutri-Score, and similar models (allergens, labels, mean/vegan/vegetarian, outliers) can be detected with similar frameworks.**
- **This means that ca 1bn humans can benefit from Nutri-Score, without brands/producers being able to prevent them!**

Models	Precision	Recall	Accuracy	F1-score
basic	0.917	0.899	0.899	0.901
using translation	0.922	0.906	0.906	0.908
with TF norm	0.919	0.903	0.903	0.905
with exp weighting	0.921	0.907	0.907	0.908
with lin weighting	0.921	0.906	0.906	0.908
with percentag w.	0.924	0.909	0.909	0.911

Table 12 Accuracy rate improvement with/without threshold

Model	With Threshold	Without Threshold	Coverage	Threshold
EN	0.979	0.972	98.7%	0.501
DE	0.955	0.923	96.0%	0.440
IT	0.964	0.944	96.4%	0.444
FR	0.942	0.914	94%	0.422

Image2Product

Can we detect products based on image or video feeds in realistic environments?

Problem: Interaction between consumers and packaged product is still limited, for example when making food choices.

- Diet tracking via diaries or mobile apps is effort-intense. **77% have never used a diet app, 18% stopped, only 5% actively use a diet tracking app**[1], despite their effectiveness when used.

- **Barcode scanning apps are not ideal for improving purchase decisions in the supermarket**, where consumers usually are not handsfree and most unhealthy food is purchased[2].

- **Computer vision-based identification of packaged products is usually done on pre-fabricated datasets and not applicable 'in-the-wild'**. Not much is known about the required setup & number of images, or achievable accuracy rates.

1) König (2018) *JMIR*, 2) Chappuis (2011), 3) Tonioni (2019)

Motivation: Make Nutrients ‚visible‘ without requiring human input

Motivation: Make Nutrients ‚visible‘ without requiring human input

Technological Advances provide New Opportunities in Passively Triggered & Hands-free User<>Product interactions

Motivation: Make Nutrients ‚visible' without requiring human input


NEU
KIND
BE-KIND, give it a try.

How can such Consumer<>Product Interaction be enabled ‘in-the-wild’, i.e. for other vending machines, supermarkets, kiosks, ... ?

Step 1: Generating Image Snippets

- Taking 300 ‘different’ pictures from the vending machine, labeling snippets

Labelled image patch

Vending machine

Step 2: Preparation of Data Set

- N=39 products for which the total labelled dataset includes at least 100 (training) + 20 (test) instances per product class.

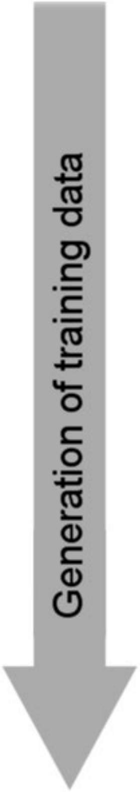
		N Classes	N Instances	%	Mean	SD
Total		109	10035	100%	92	129
	Beverages	51	5646	56%	110	163
	Snacks	58	4389	44%	76	86

- In order to support research on computer vision-based product detection, we invite interested researchers to find and use our labelled dataset:
  - https://www.autoidlabs.ch/projects/holoselecta/
  - Please cite if you use it: Fuchs, K., Grundmann, T., Fleisch, E., **Towards Identification of Packaged Products via Computer Vision**, The 9th International Conference on the Internet of Things (IoT 2019)

Step 3: Set up CV pipeline

- Generating training & test data dynamically
- Calibrating the neural networks
  - Replacing only last layer
  - Retraining with very small learning rate
  - Manual tuning
  - Google Cloud GPU
- Object detection and Image classification

1. **Inception** V2 for IC, with Faster RCNN for OD
2. **Resnet** 50 V2 for IC, with SSD and Focal Pyramid Networks (Retinanet) for OD
3. **Mobilenet** V2 for IC, with SSD for OD

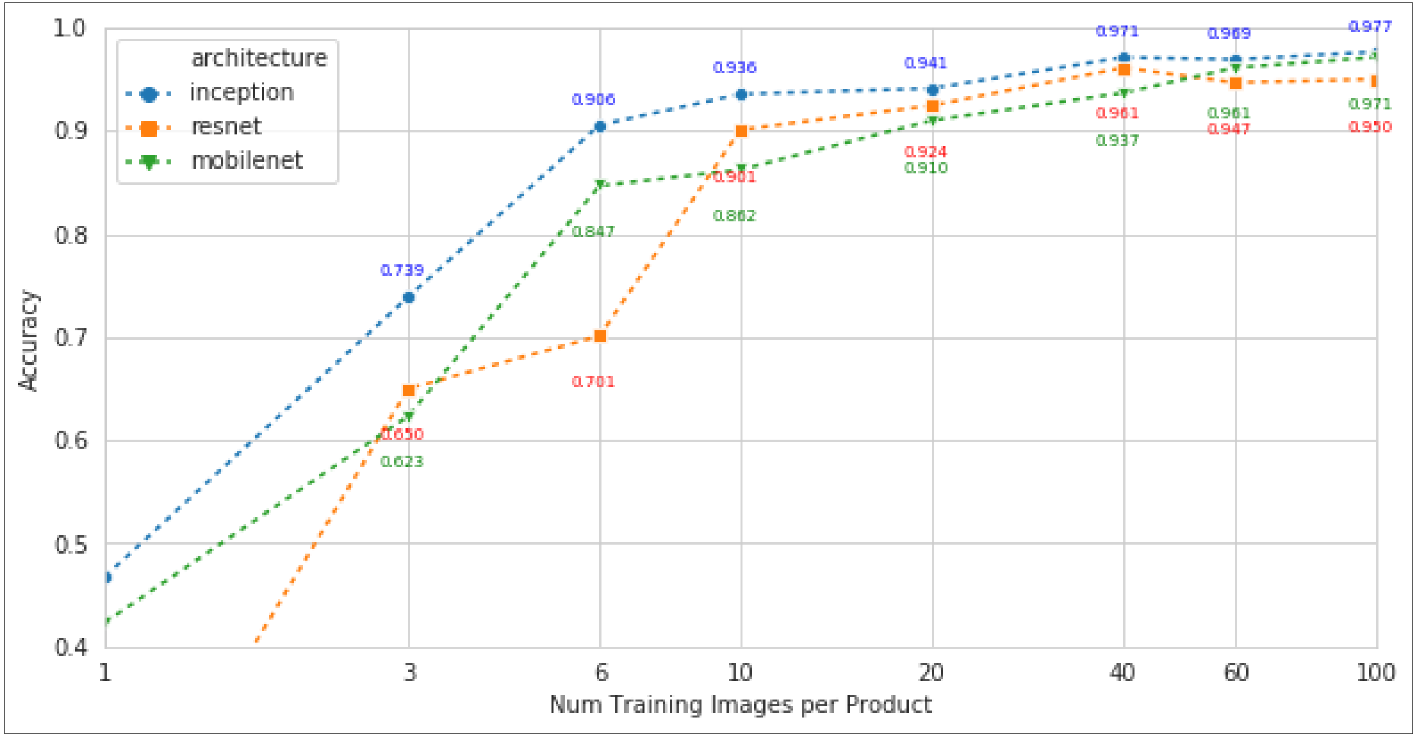


300 labelled images of vending machine assortment

Training data with k labelled images patches per product class (not needed images blacked out)

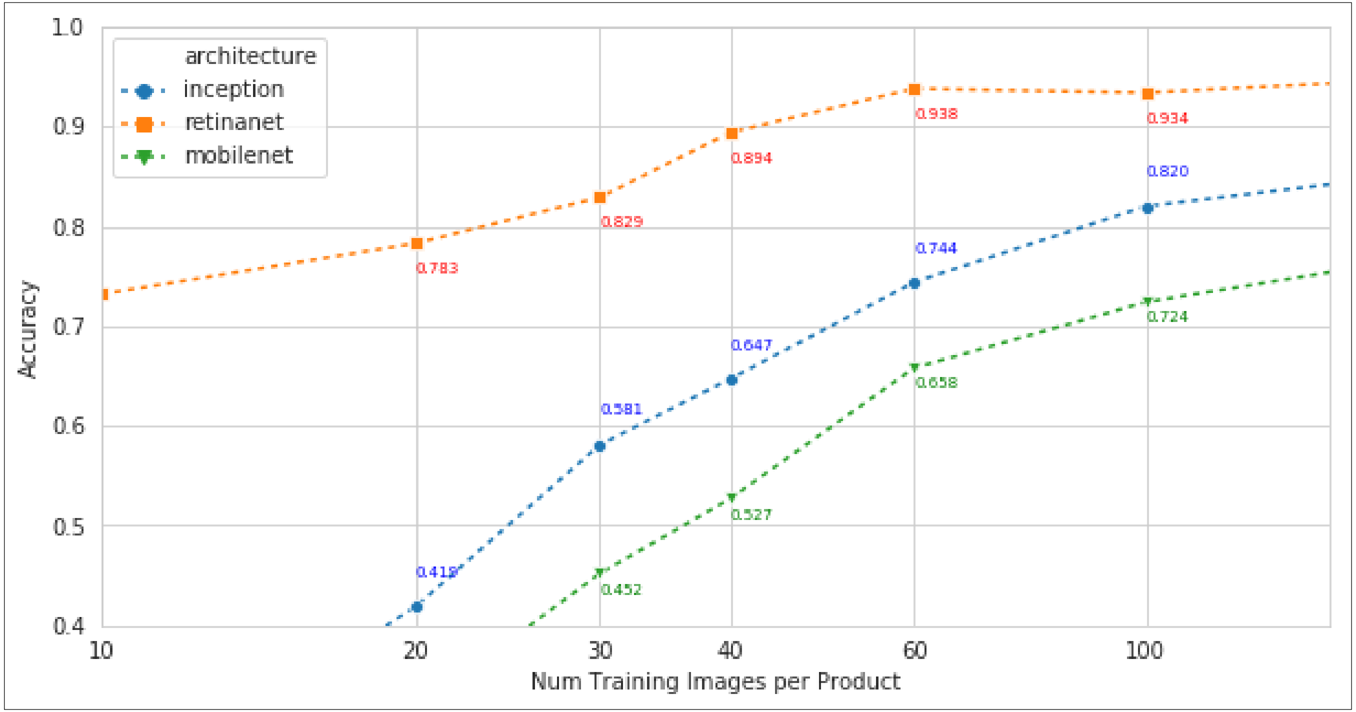
Image Classification (N=39 classes)

- Only 6 images for 90% Accuracy
- At least 26 images for 95% Accuracy
- At least 100 images for 97.5%



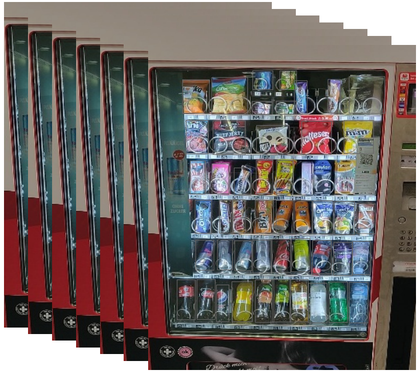
Object Detection (N=39 classes)

- Mean average precision (mAP) for intersect over union (IoU) of 0.5
- Only Resnet achieved mAP 75% over 0.9 with at least 42 images
- Architecture choice very relevant

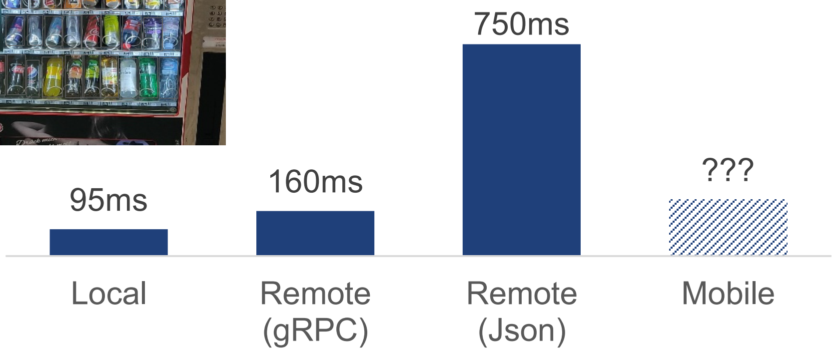


Never 100% accuracy, but image pooling allows for improved accuracy

- Choice of neural network, connectivity, image resolution, hardware capability influence how many frames per second can be assessed
- For multiple frames per second, accuracy increase via image pooling (until edge cases remain)



1 sec @ 7 fps



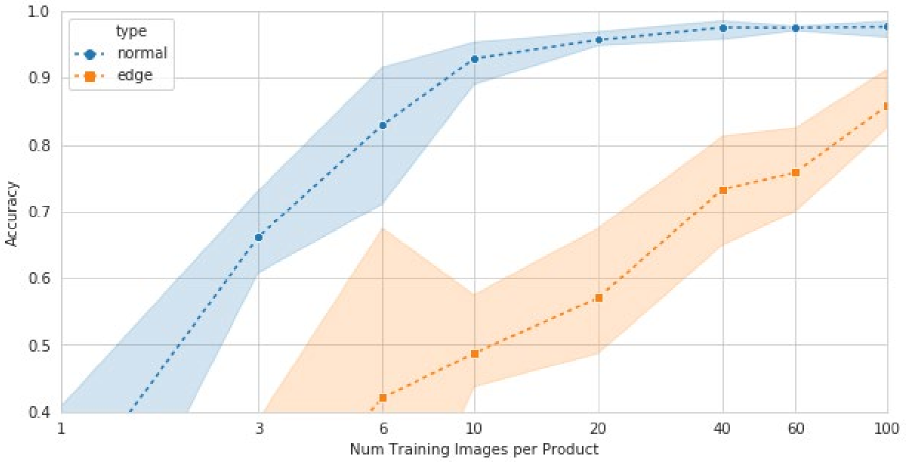
$$1 - (0.2)^7 = 0.99999$$

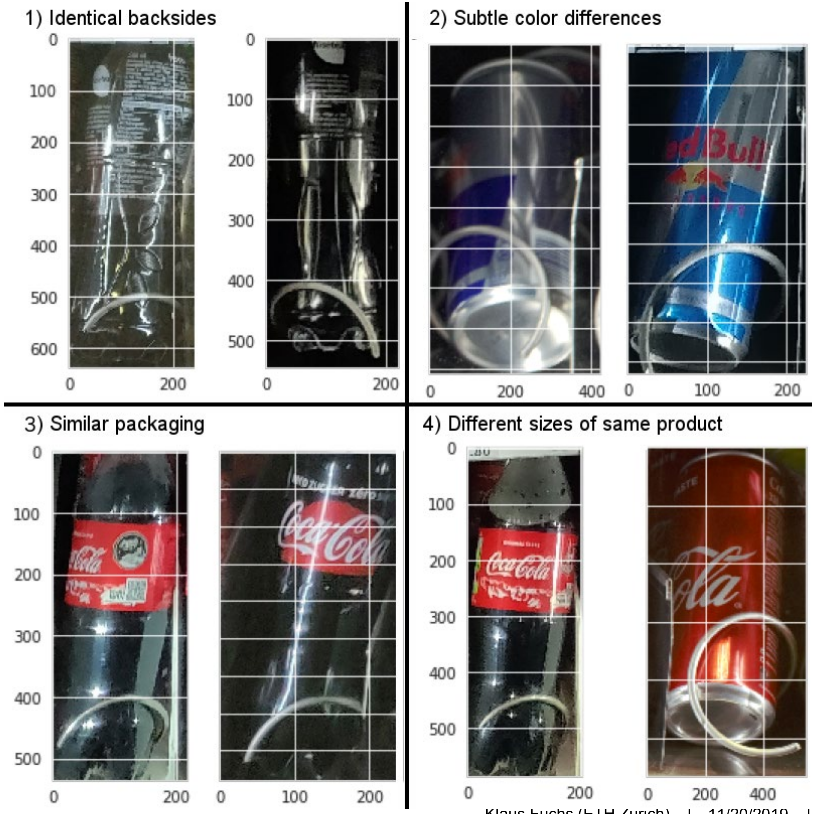
1 in 100000 fails for 1 sec

Border cases heavily impact the error rate

$$a_{max} = 98.5\%$$

→ almost half the error rate





Conclusion: Computer vision has advantages but also disadvantages over barcode scanning.

✚ Pro

- Gives relative Position of user to object
- Detection of multiple objects at once
- Dimensions of objects retrievable
- Passively triggered

━ Contra

- Technical complexity of models
- Image resolution
- Amount of images needed
- Computational requirements (GPUs)
- Never 100% accuracy, edge cases exist

Conclusion: Machine learning & AI are coming, here are my tipps:

- **Make sure your data allows real-world AI application:** Check your labeled data, data quality, feature variety before starting your AI / ML task.
- **Computer vision-based product identification works at acceptable accuracy for non-critical use-cases** (e.g. labels, advertisements, showing recipes, ingredient information, store planning, inventory, but not yet for self-checkout)
- **Limitations lie in the image capturing process**, **one-time setup** (product packages change ca every 2 years), **generalizability** not guaranteed
- **Global trade item number (GTIN) as identifier remains relevant**, also to find and link other information (Digital Link)
- **Knowledge Graph needed** (Future work: Location, Retailer, Shelve > reduce number of possible classes from millions to dozens)
- Computer vision on packaged products **requires much more labeled data from the real world** to support text extraction (OCR), brand extraction

Contact

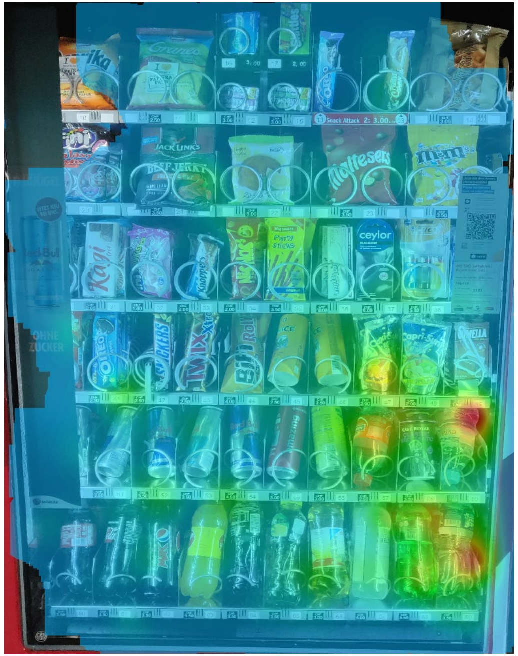
- Follow our research: www.autoidlabs.ch
- Follow my research: https://www.researchgate.net/profile/Klaus_Fuchs2
- Klaus L. Fuchs
  Associate Research Director
  Auto-ID Labs ETH Zurich / HSG

- ETH Zurich WEV G219
  Information Management
  Department of Management, Technology, and Economics
  Weinbergstrasse 56-58
  8092 Zurich

  http://www.autoidlabs.ch
  Phone +41 78 858 70 37

Discussion & Thank you

Consumers spend different amount of time exploring the content of the machine, applications could use that information

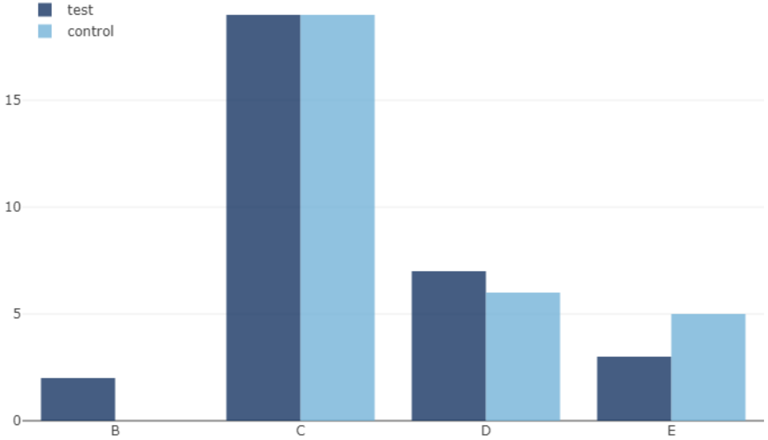



NUTRI-SCORE
A B C D E
Healthy
Unhealthy
Selecta App:
Get your discount now!
Snack Attack
PAY WITH ...

Nutri-Score visualization (Test) changes selection of beverages!

Snacks

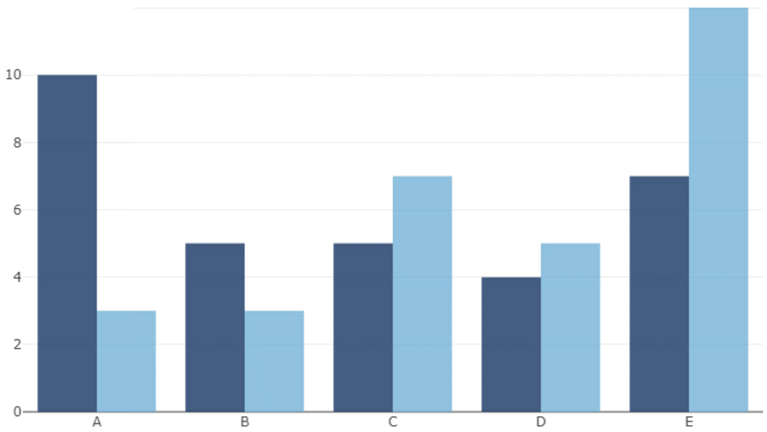
	Mean	(SD)
Test	10.00	(4.71)
Control	11.37	(5.42)



Buying a snack of their choice ($p = 0.3502$)

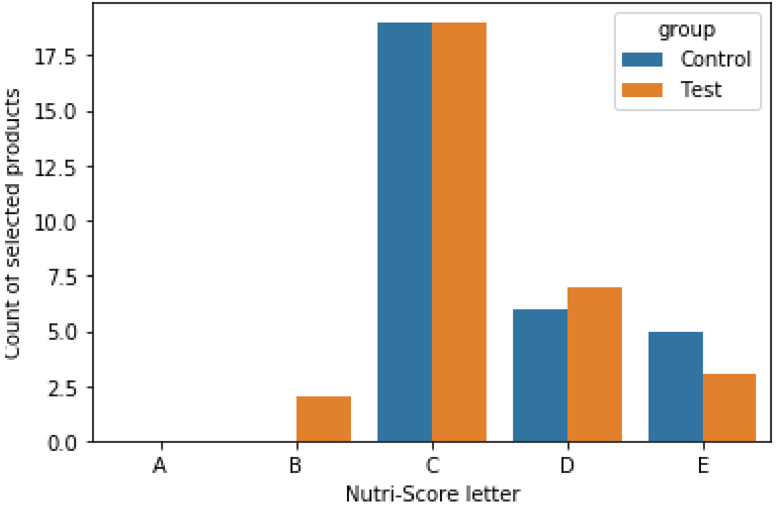
Drinks

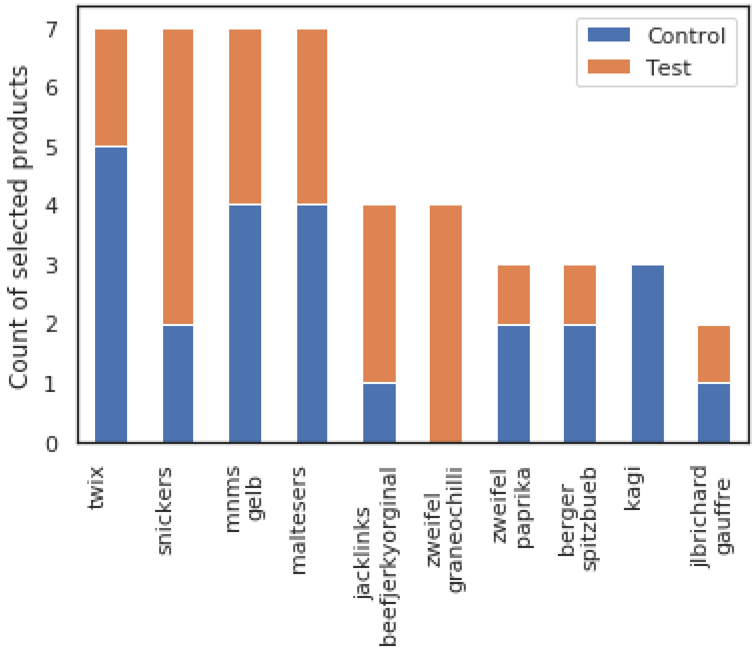
	Mean	(SD)
Test	-0.97	(10.33)
Control	4.80	(7.59)

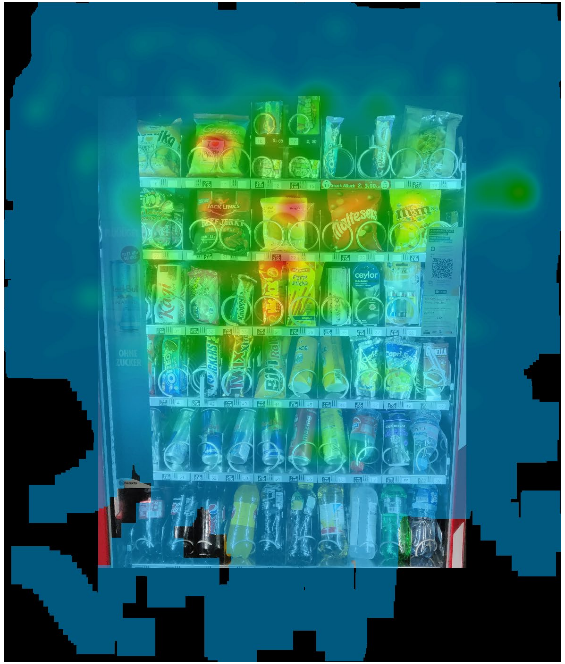


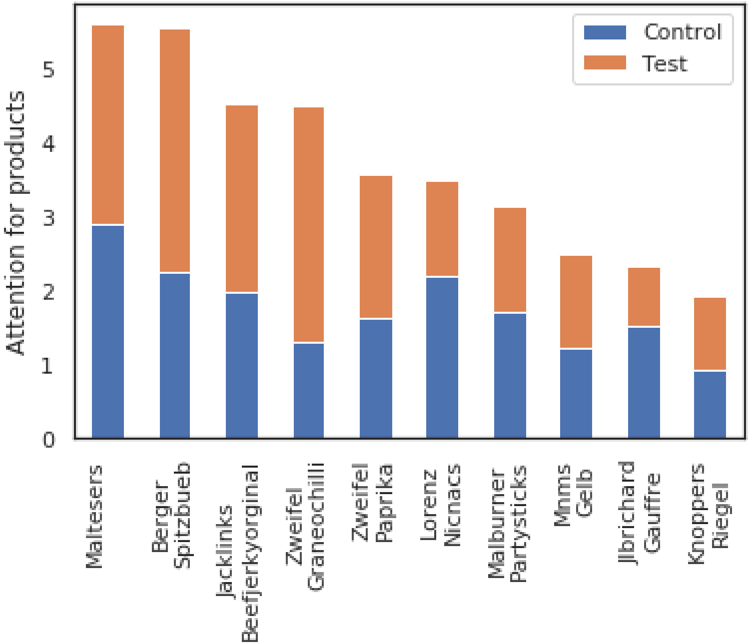
Buying a drink of their choice ($p = 0.0092$)

Results Task 1: Select a snack of your choice.

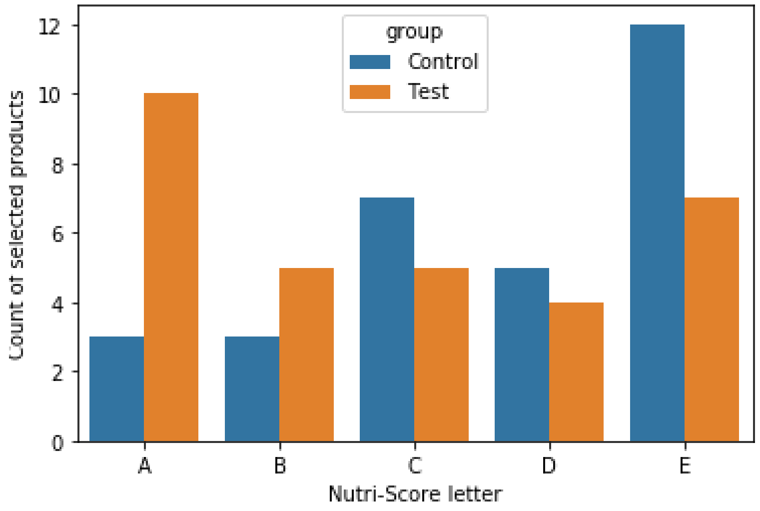


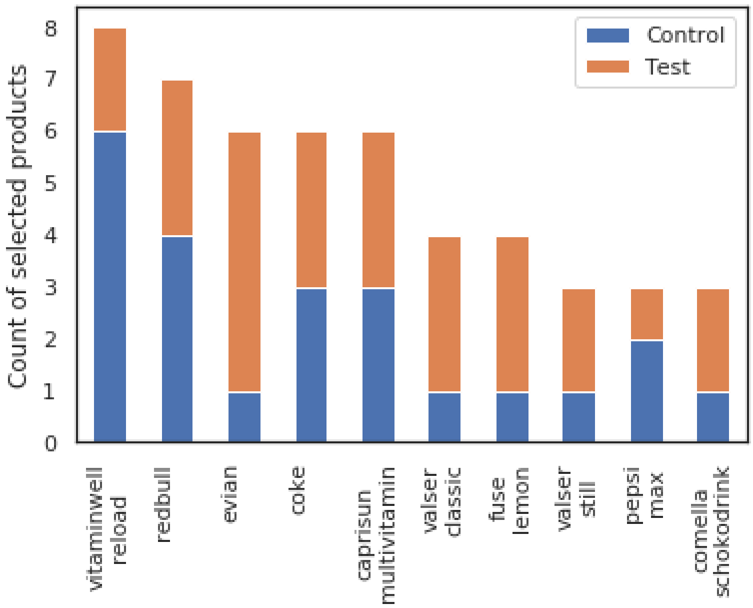


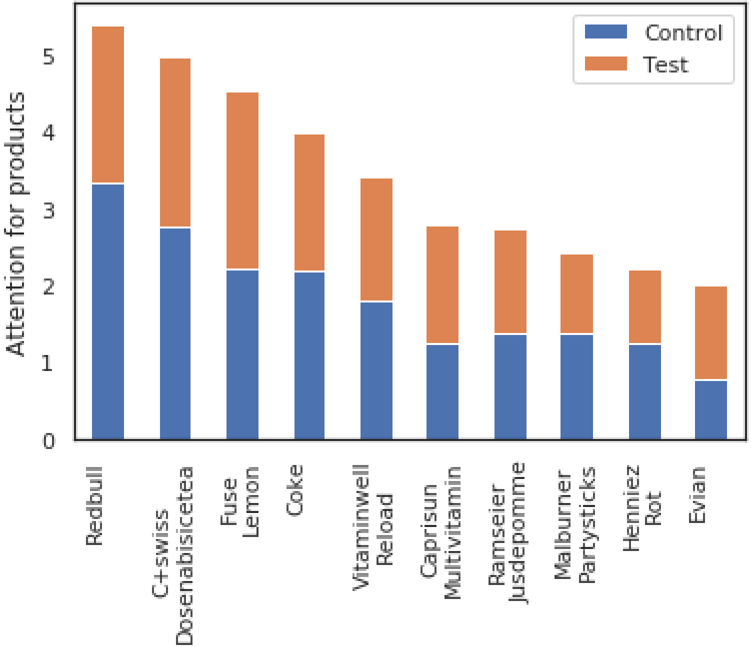



Control
Test
Attention for products
0
1
2
3
4
5
Maltesers
Berger Spitzbueb
Jacklinks Beefjerkyorginal
Zweifel Graneochilli
Zweifel Paprika
Lorenz Nicnacs
Malburner Partysticks
Mnms Gelb
Jlbrichard Gauffre
Knoppers Riegel

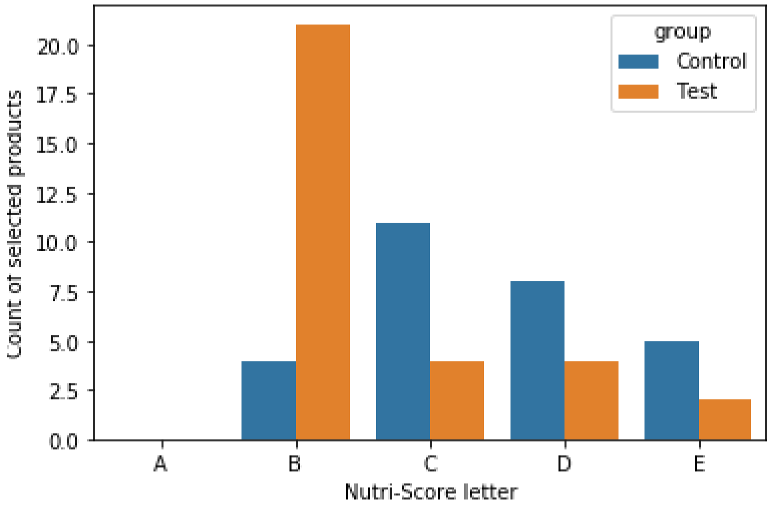
Results Task 2: Select a drink of your choice.

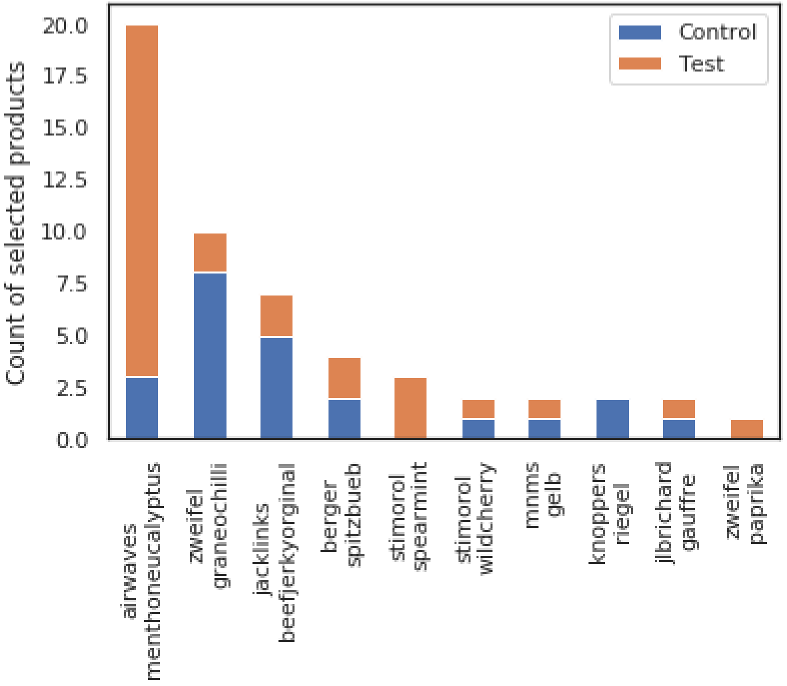


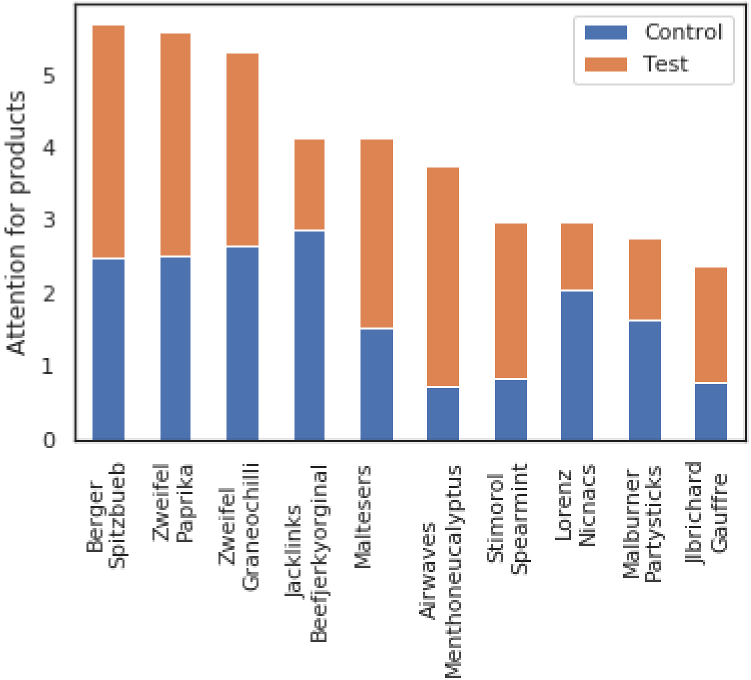



Control
Test
Attention for products
0
1
2
3
4
5
Redbull
C+swiss Dosenabisicetea
Fuse Lemon
Coke
Vitaminwell Reload
Caprisun Multivitamin
Ramseier Jusdepomme
Malburner Partysticks
Henniez Rot
Evian

Results Task 3: Select the healthiest snack.






Control
Test
Attention for products
0
1
2
3
4
5
Berger Spitzbueb
Zweifel Paprika
Zweifel Graneochilli
Jacklinks Beefjerkyorginal
Maltesers
Airwaves Mentheucalyptus
Stimorol Spearmint
Lorenz Nicnacs
Malburner Partysticks
Jlbrichard Gauffre

Results Task 4: Select the healthiest drink.