



Deep learning for retail analytics and reference data management

Alessandro Zolla
Robert Bogucki

Nielsen measures what people...

WATCH

- TV Ratings
- Advertising exposure
- TV and Digital media

BUY

- Brick & Mortar
- eCommerce
- FMCG

100+ countries

40,000+ employees

10M+ active products

Nielsen Reference Data: industry standard for analytics

What is Reference Data?

It's the glue that brings Nielsen's assets together, enabling internal and external data exchange.



Our Strategy:

1. Create **Foundational Content**, leveraging internal resources and partners
2. Build normalized layer of **Analytic Ready** content
3. Deploy **automation** to deliver faster and with quality
4. Enable content ecosystem and data exchange

Nielsen RD Layered Content



Layered Reference Data



Market Behavior Dynamic Chars

- Dynamic Characteristics based on market place data
- e.g. On-Line only, Purchasing Demographic based



Client Maintained Characteristics

- Characteristics are fully created, coded and maintained by Client



Innovation

- Characteristics are managed and maintained by Nielsen
- Dynamically maintained from Analytical Ready and Foundational Characteristics



Health & Wellness

- Characteristics are managed and Maintained by Nielsen or Nielsen Partners
- Utilize Analytical Ready and Foundational Characteristics
- Can cover H&W, Sustainability, Ethical Sourcing, etc.



Client Ready Content

- Characteristics are created by mapping rules by Nielsen following Client Definition
- Utilize Analytical Ready
- May include Client Custom views of H&W, Innovation, Analytical Ready, etc.



Analytical Ready

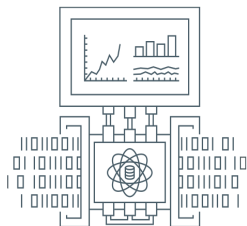
- Universal and Category relevant Characteristics identified and designed by Nielsen
- Harmonized, dictionary based, values consistent & ready for use



Foundational Characteristics

- All pack specific information included, i.e. Ingredients, nutrition panel, claim
- Pack in Hand/Picture based coding required
- Unstructured and not dictionary managed

Why Deep Learning?

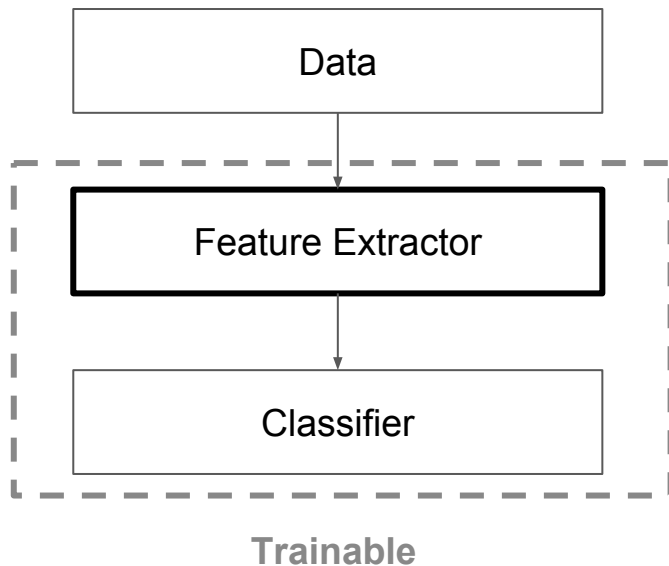


Machine & Deep Learning is extracting knowledge from data

- no need to know how to solve the problem to solve it
- works with all sort of data (text, images, signals and more)
- similar techniques viable across many problems and sectors

Why Deep Learning?

Deep Learning



Fully Trainable Model:

- End-to-end learning
- Self-generated high-level features
- Fine-tuned to your problem

What's on the package?

nielsen

deep**sense**.io
BIG DATA SCIENCE

Things you may be interested in:

- Barcode
- Brand logo
- Nutritional facts
- Ingredients
- Size
- Recycling information
- Allergy advice
- Producer information
-



What's on the package?

nielsen

deep**sense**.io
BIG DATA SCIENCE

Things you may be interested in:

- Barcode
- Brand logo
- Nutritional facts
- **Ingredients**
- Size
- Recycling information
- Allergy advice
- Producer information
-



Case Study: Ingredients

nielsen

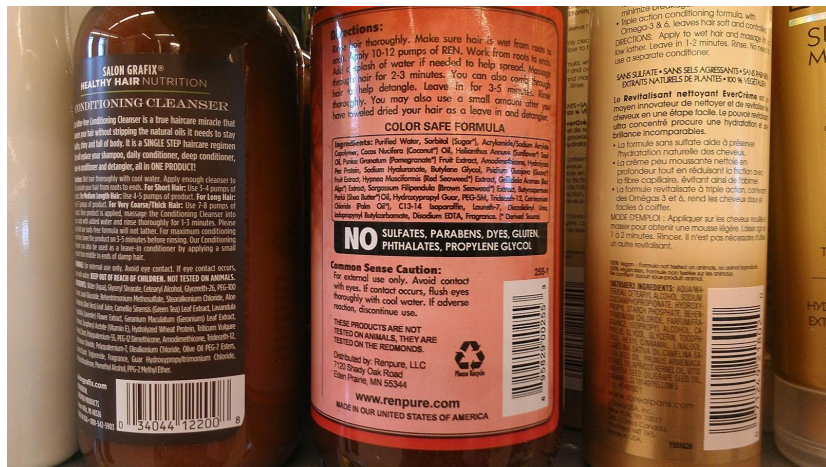
deep**sense**.io
BIG DATA SCIENCE

Problem:

Find the region containing the **ingredients** of the product images

Challenges:

- Reflections
- Bends
- Foil
- Close to impossible without understanding the text
- ...



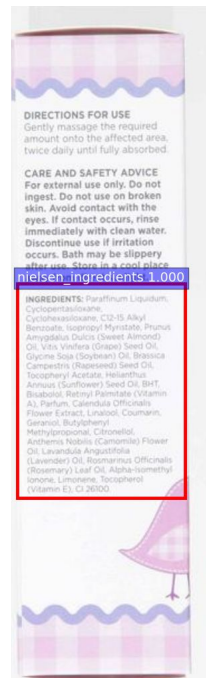
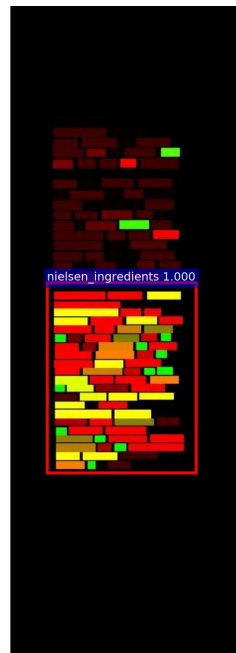
Case Study: Ingredients

nielsen
.....

deep**sense**.io
BIG DATA SCIENCE

Feature engineering:

- Heatmap of ingredients-like words
- Commas
- The word “Ingredients”



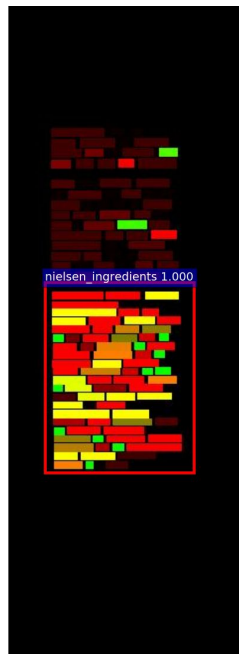
Case Study: Ingredients

Feature engineering:

- Heatmap of ingredients-like words
- Commas
- The word “Ingredients”

Simple heuristics:

- A decent sized rectangular shape with many blobs inside
- A decent sized rectangular shape starting from the “ingredient blob”...



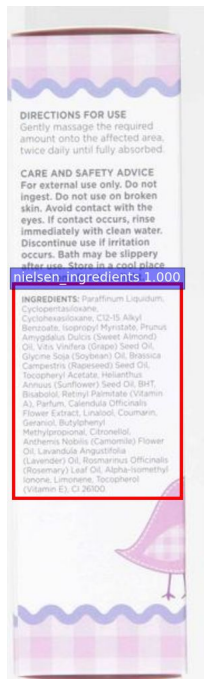
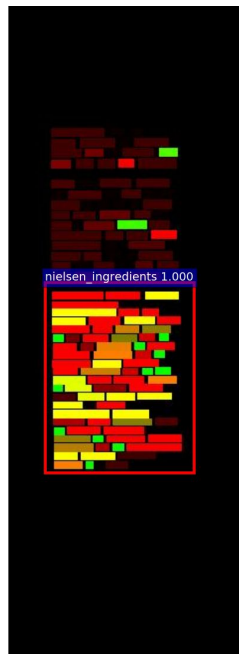
Case Study: Ingredients

nielsen
.....

deep**sense**.io
BIG DATA SCIENCE

We need to go deeper:

- Original image gives us a good feeling where the area is, but we may not be able to decide without reading the words



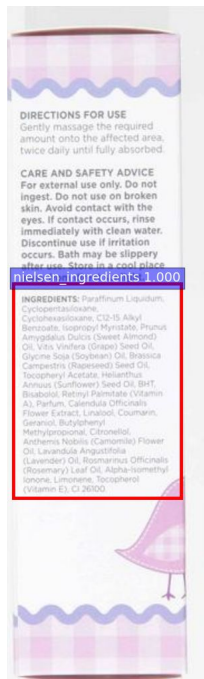
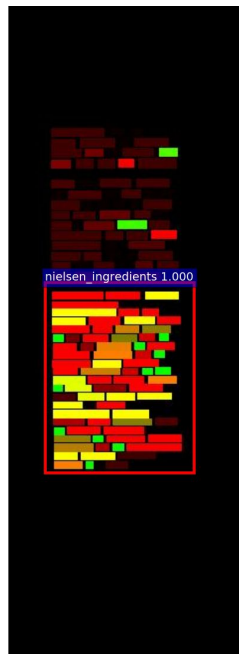
Case Study: Ingredients

nielsen

deep**sense**.io
BIG DATA SCIENCE

We need to go deeper:

- Original image gives us a good feeling where the area is, but we may not be able to decide without reading the words
- Heatmaps give us the way to understand the content, but they ignore the visual information



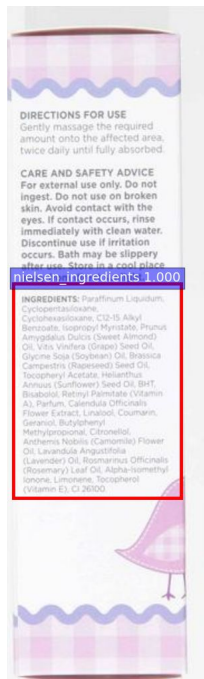
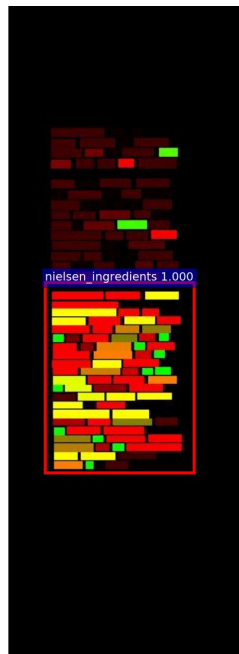
Case Study: Ingredients

nielsen

deep**sense**.io
BIG DATA SCIENCE

We need to go deeper:

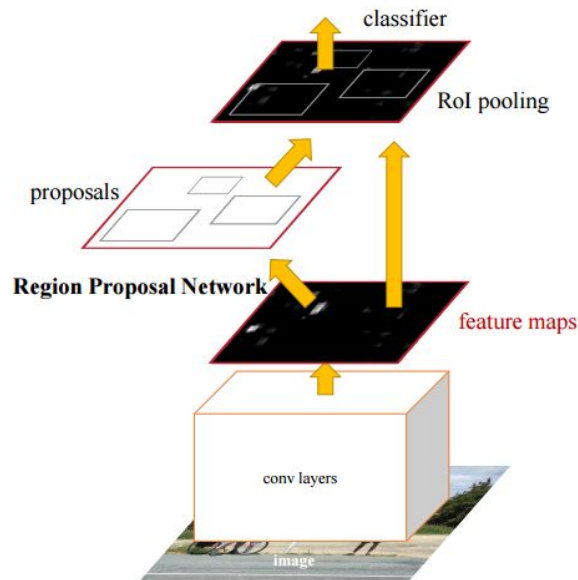
- Original image gives us a good feeling where the area is, but we may not be able to decide without reading the words
- Heatmaps give us the way to understand the content, but they ignore the visual information
- But it's easy to have both with deep learning!



Case Study: Ingredients

Faster RCNN:

- State of the art object detection network
- Region Proposal Network:
“where to look”
- Detector Network:
“what do I see”
- Both networks use the same feature maps
- Based on VGG-16



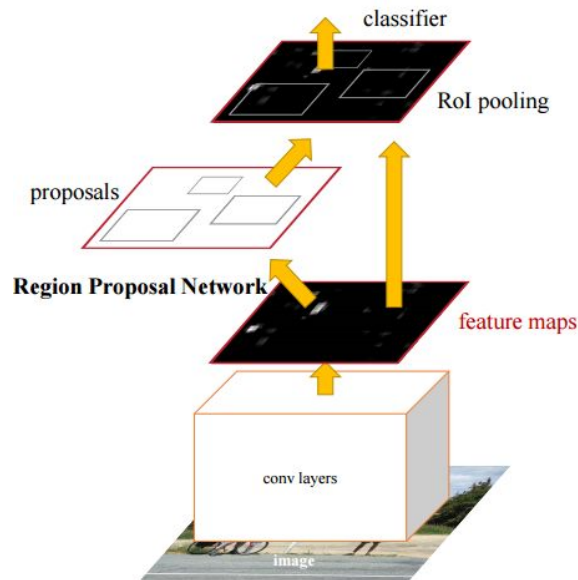
Case Study: Ingredients

Final solution in a nutshell:

- Use original image input
- Add text-based additional features as images on different channels
- Run Faster RCNN

Outcome:

- Over 90% accuracy



Nourished care shampoo

Hydrate and moisturise dry or damaged hair.
Nourished care shampoo by Sainsbury's
morning oil, known for its rich moisture,
help you maintain nourished, supple skin
to natural vitality.

Directions for use
For best results use with Nourished care conditioner.
Apply a generous amount of shampoo to wet hair.
Gently massage into your roots and work down the
length of the hair. Rinse thoroughly. Repeat if required.

Safety

nilsen ingredients 1000

Our ingredients: Aqua, Sodium Lauryl Sulphate, Cocamidopropyl Betaine, Glycol Distearate, Potassium Sorbate, Disodium EDTA, PEG-9 Glyceryl Ether, Sodium Polyacrylate, PEG-8 Diethylamine, PEG-6 Dimethylamine, Quaternium-15, Parfum, Sodium Hydroxide, Hexyl Cinnamal, Linalool, Benzyl Alcohol, Propylene Glycol, Phenoxyethanol, PEG-30 Castor Oil, Vitamin C, Citric Acid, Magnesium Nitrate, Isopropyl Alcohol, Methylparaben, Butylparaben, Ethylparaben, Glycerol, Citrus Acid, Borage Seed Oil, Hippocastanum Seed Extract, Potassium Dihydrogen Phosphate, Sodium Benzoate, Sodium Chloride, Sodium Hydroxide, Calcium Panthenol, Methylchloroisothiazolinone, Methylisothiazolinone, Bisphenol A.

BOTTLE - PLASTIC
worry recycled

CAP - PLASTIC
check local recycling

Produced in the UK for
Sainsbury's Supermarkets Ltd, London EC9B 2HT
Coveline (P)02 1826761
www.sainsburys.co.uk

24M **Crabtree & Wey**

[illegible]

nielsen ingredients detections



Some examples

nielsen

deep**sense**.io
BIG DATA SCIENCE

nielsenIngredients detections



nielsenIngredients detections



- A data-analytics brand by **CodiLime** - ranked 2nd in Deloitte CE 2016 Technology Fast 50 list
- **200** people on board in two locations - Poland and California
 - > 120 Software Engineers, > **40 Data Scientists** and growing
 - Winners at Kaggle & various algorithmic competitions
- Providing machine and **deep learning** solutions and consultancy
- Working with market leaders, such as:



Thank you for your attention