# Deep Context for Fine-Grained Object Recognition in Crowded Images

ERAN GOLDMAN

Joint work with Prof. Jacob Goldberger,

Faculty of Engineering, BIU

Trax image recognition

# AGENDA

- Trax Visual Challenges
- Deep Context Embedding Architecture
- Implementation Details
- The Detection Challenge
- Summary

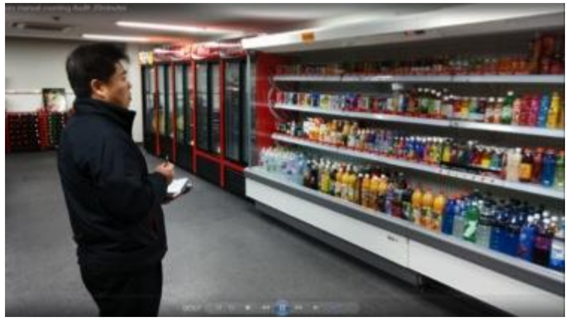



Trax image recognition

# Traxs' Business Application

Manual Audit

Trax Automatic Recognition Audit



Slow and Expensive
Inconsistent
untraceable

Fast and Cheap
Consistent
Traceable

**'Big Data' for retail**

AVAILABILITY

SHARE OF SHELF

PRICING


PROMOTIONAL ACTIVATIONS

COMPETITVE INSIGHTS

PLANOGRAM COMPLIANCE

SHELF STANDARDS




Trax image recognition

# Trax unlocks ‘Big Data’ for the retail industry

## Scale of the Data

### Market
Market share, trade channels, segments, competition

### Retailer
Brand champions, generic brands, range review

### Store
Product location, assortments, shelf share

### Shelf
NPD, POSM, Pricing, Promotions

## Scale of coverage




Trax image recognition

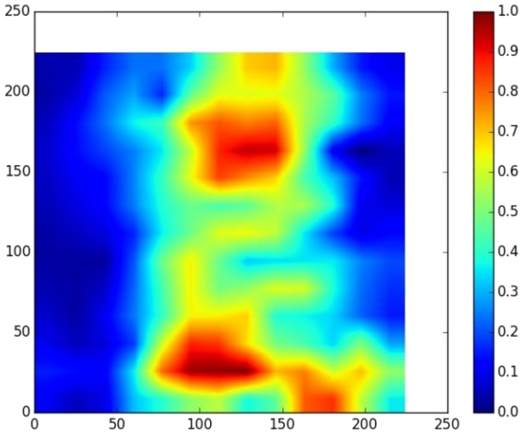
# Welcome to Trax Universe

# Trax's Visual Challenges

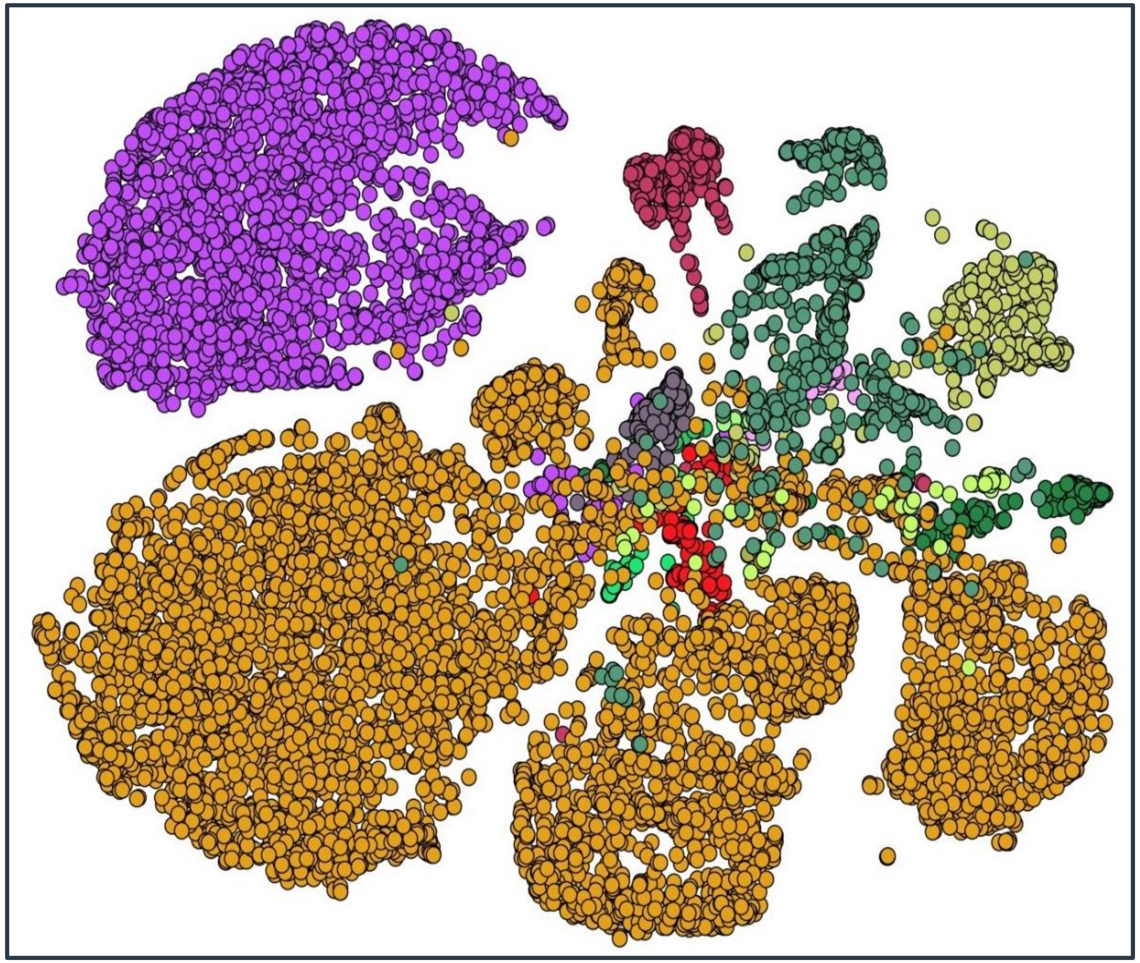
- # Classes
- Fine-Grained Classification
- Crowded Scene
- Dynamic Dataset

# Last Year- CNN: Opening The Black Box





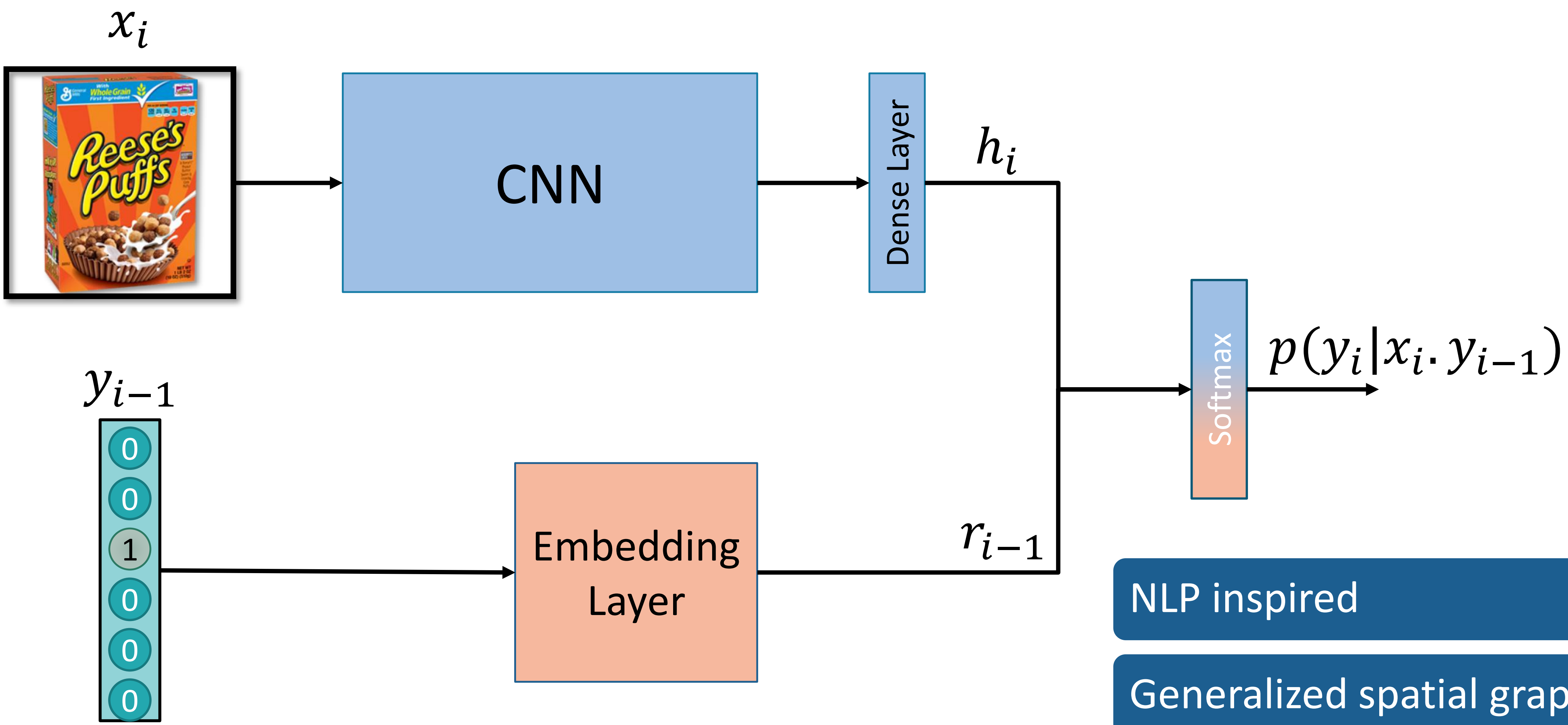
# AGENDA

- Trax Visual Challenges
- Deep Context Embedding Architecture
- Implementation Details
- The Detection Challenge
- Summary



Trax image recognition

# New DNN architecture

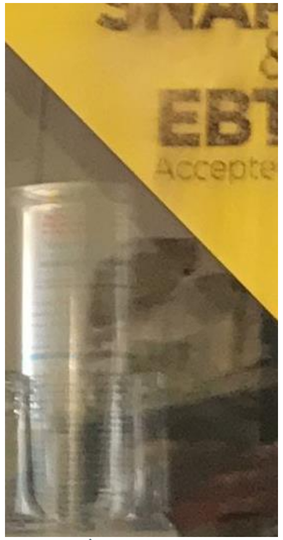



NLP inspired

Generalized spatial graph

# Which one is it?

classes




Object

# Which one is it?

classes



Object

# Which one is it?

classes



Object

# Which one is it?

classes



Object

# Which one is it?

Object

classes

Flavors with red labels:

| Classic | Decaf | Lemon | Vanilla | Ginger |
|---|---|---|---|---|

X Sizes:

| 0.35 | 0.47 | 0.5 | 0.6 | 1 | 1.25 | 2 |
|---|---|---|---|---|---|---|




Trax image recognition

# Which one is it?

Object


classes

Flavors with red labels:

| Classic | Decaf | Lemon | Vanilla | Ginger |
| --- | --- | --- | --- | --- |

X Sizes:

| 0.35 | 0.47 | 0.5 | 0.6 | 1 | 1.25 | 2 |
| --- | --- | --- | --- | --- | --- | --- |




Trax image recognition

# Which one is it?

Object


classes

| | | | | | | |
|---|---|---|---|---|---|---|
| 0.35 ltr (12 OZ) | 0.47 ltr (16 OZ) | 0.5 ltr | 0.6 ltr (20 OZ) | 1 ltr | 1.25 ltr | 2 ltr |




Trax image recognition

# Which one is it?

Object

classes

| 0.35 ltr (12 OZ) | 0.47 ltr (16 OZ) | 0.5 ltr | 0.6 ltr (20 OZ) | 1 ltr | 1.25 ltr | 2 ltr |
|---|---|---|---|---|---|---|





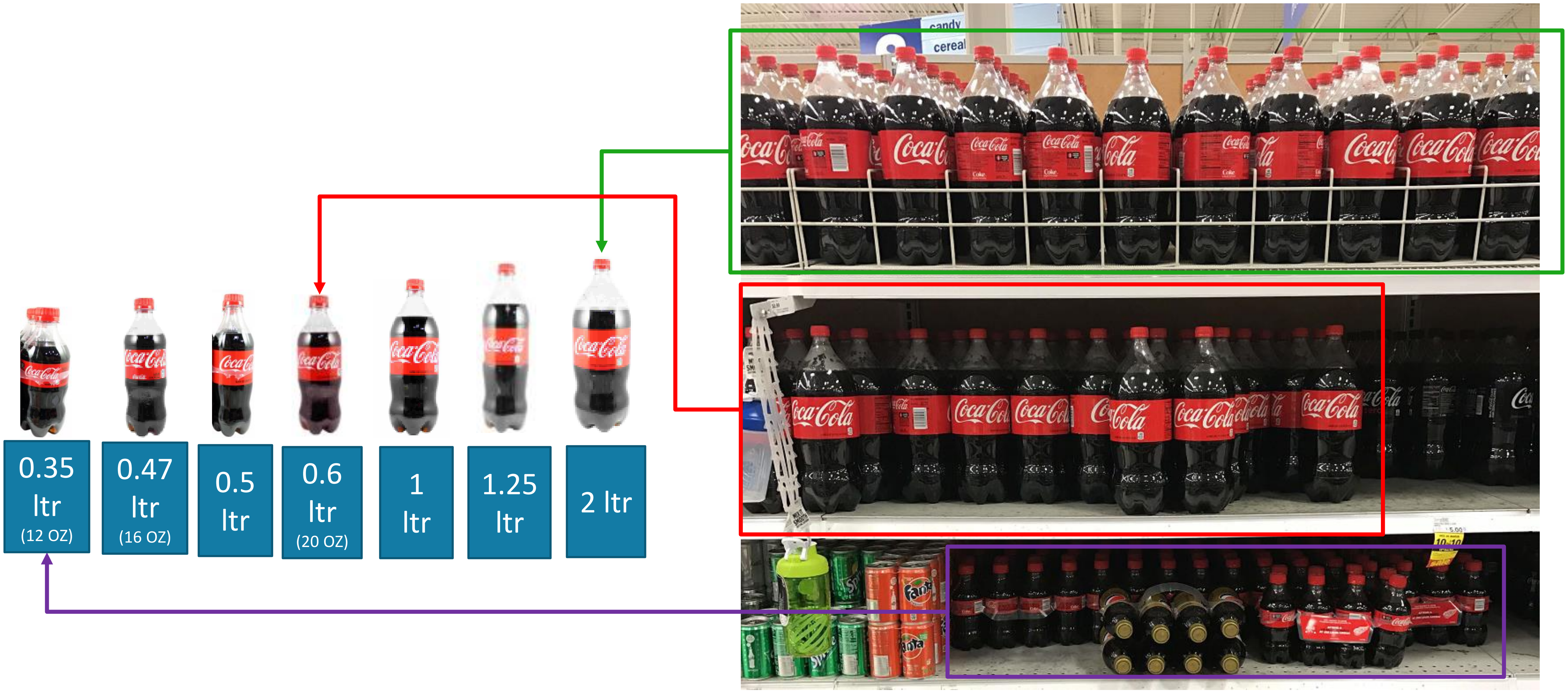
Trax image recognition

# Which one is it?

Object


classes


| 0.35 ltr (12 OZ) | 0.47 ltr (16 OZ) | 0.5 ltr | 0.6 ltr (20 OZ) | 1 ltr | 1.25 ltr | 2 ltr |
| --- | --- | --- | --- | --- | --- | --- |


0.35 ltr (12 OZ)
0.47 ltr (16 OZ)
0.5 ltr
0.6 ltr (20 OZ)
1 ltr
1.25 ltr
2 ltr
candy
cereal

# Natural Language Processing Analogy

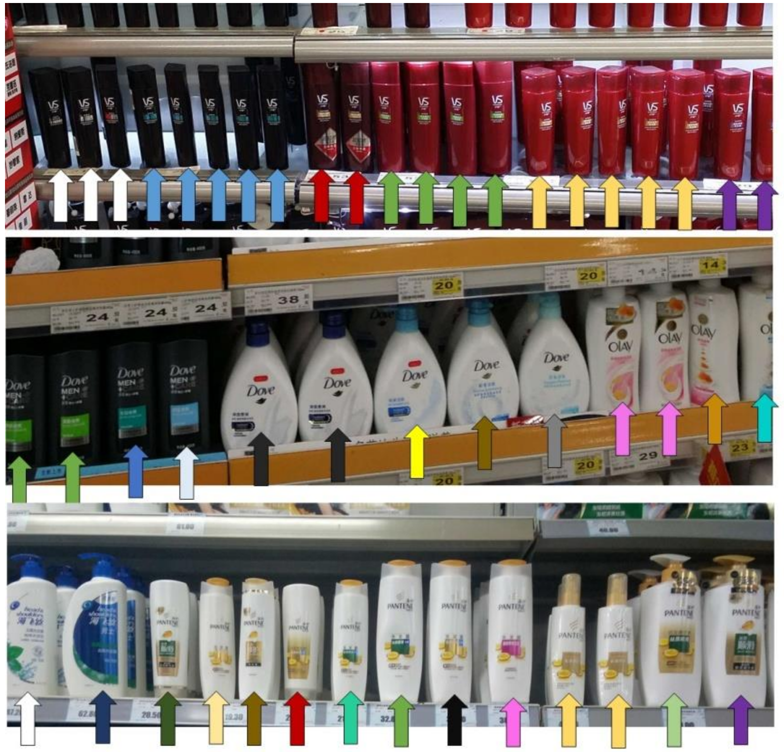
SENTENCES

SHELVES

A tall glass of orange juice is the very image of refreshment.

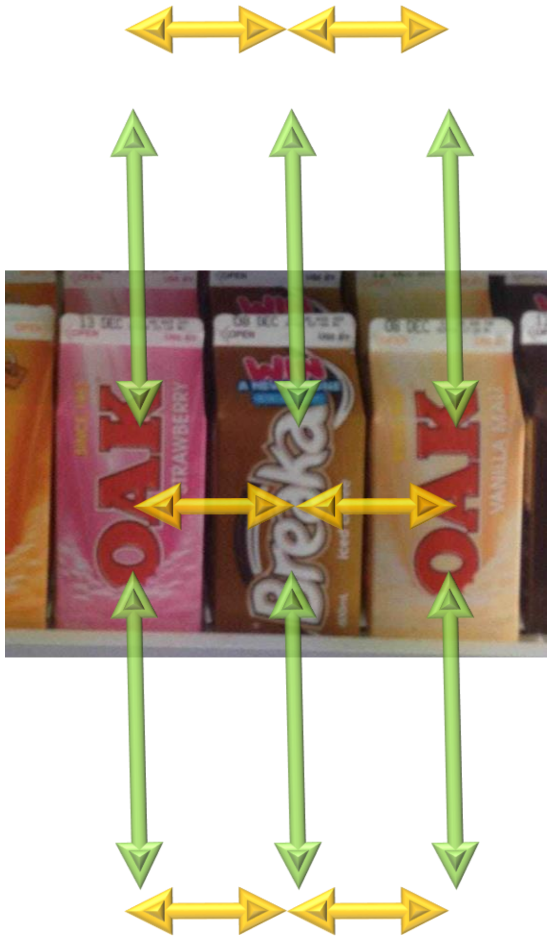
Overwatering an orange tree can cause the leaves to turn pale.

The fiber content in an orange peel makes you feel fuller after eating.






Trax image recognition

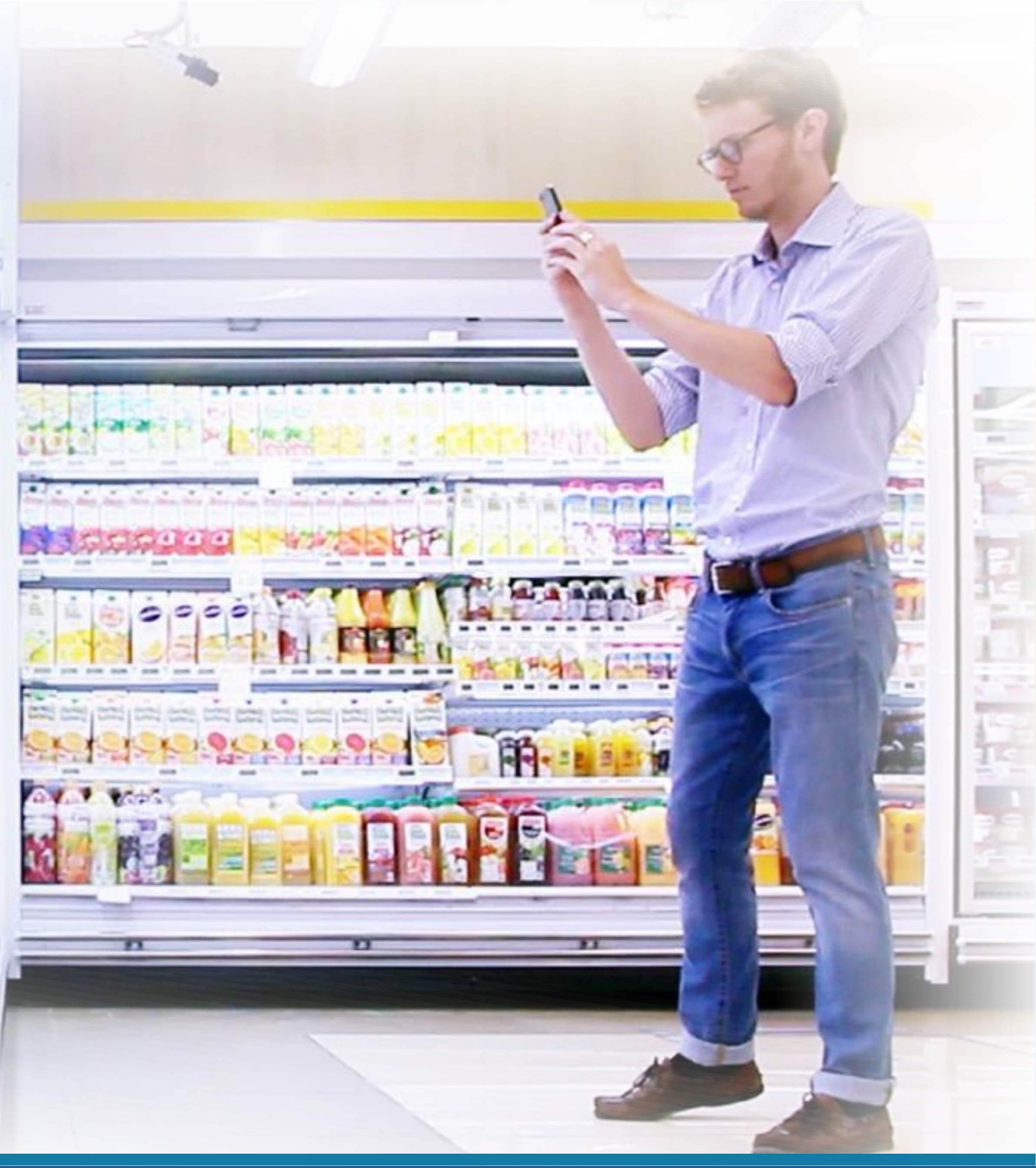



Yes We Can

# AGENDA

- Trax Visual Challenges
- Deep Context Embedding Architecture
- Implementation Details
- The Detection Challenge
- Summary





Trax image recognition

# Extended "Sentence"

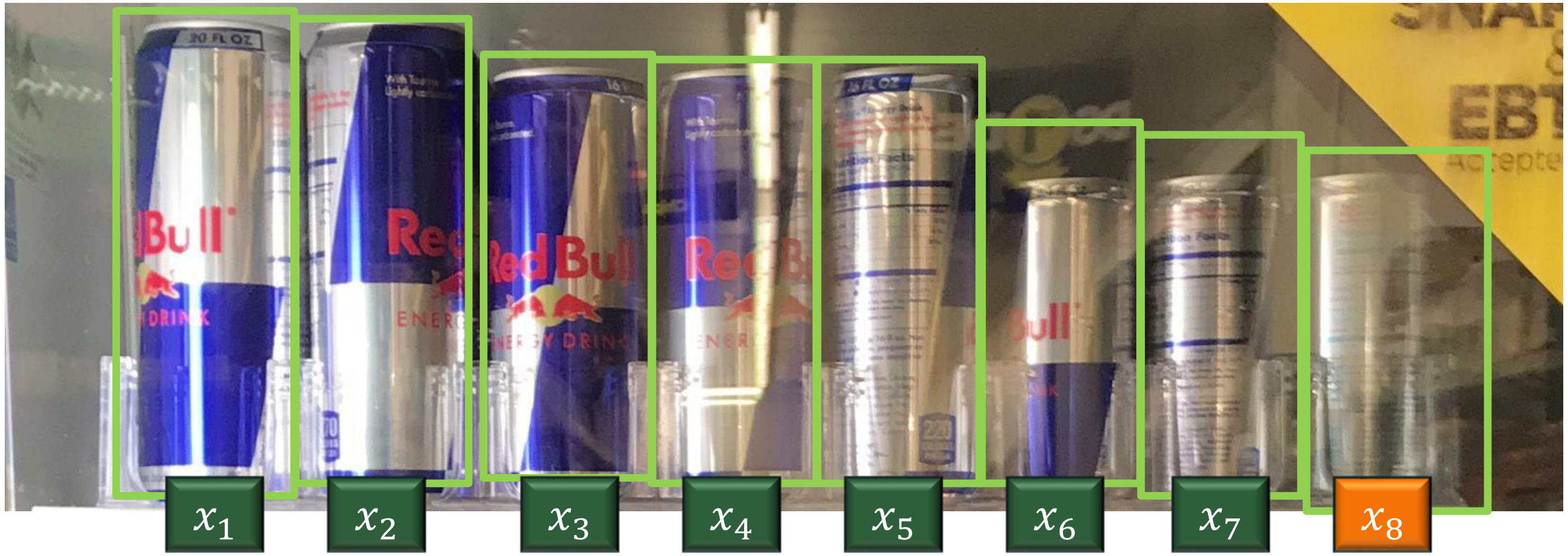
## Model image as a graph

- Detection Node
- Left-Right Relation Edge
- Top-Bottom Relation edge

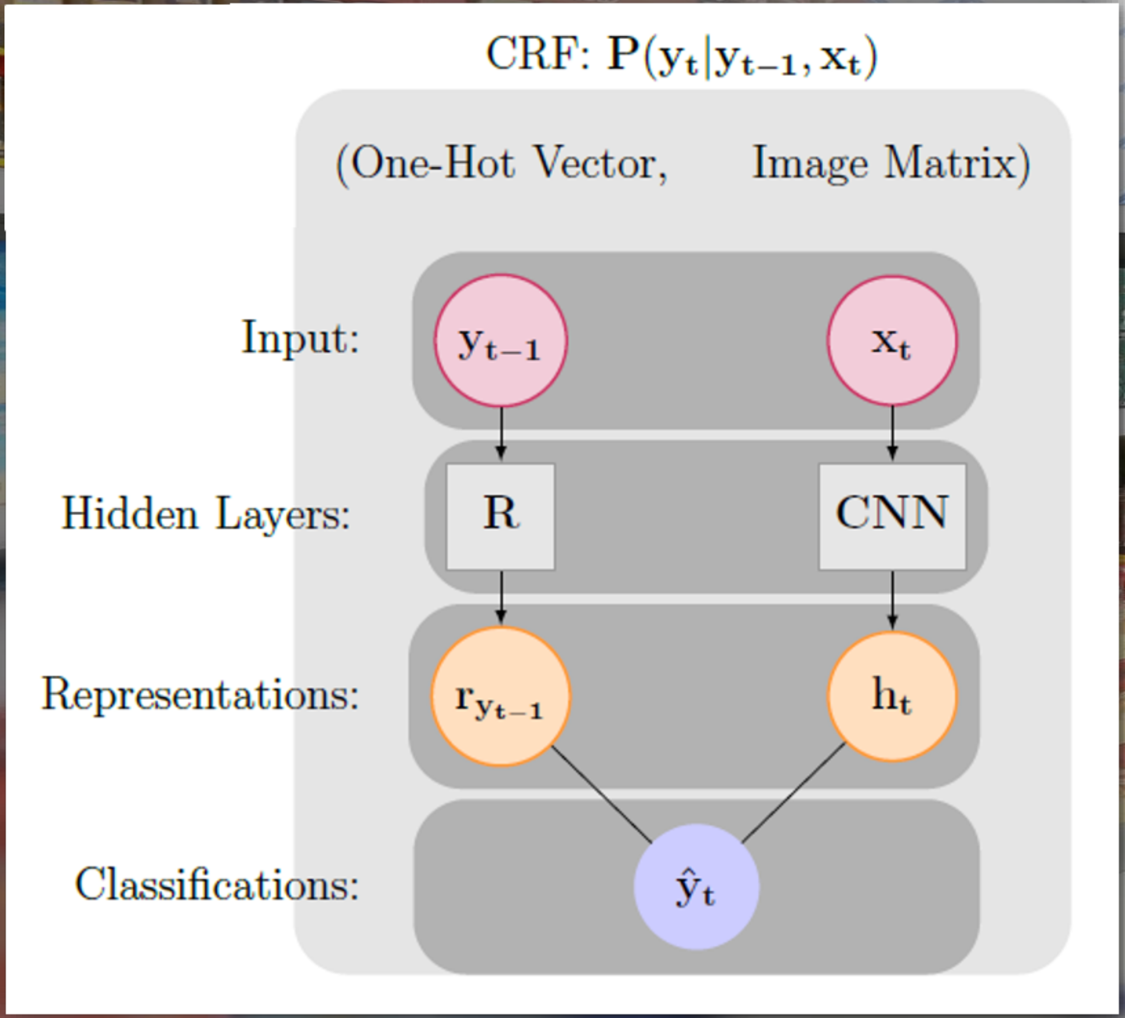
# Classification


$$P(y_8|x_8)$$

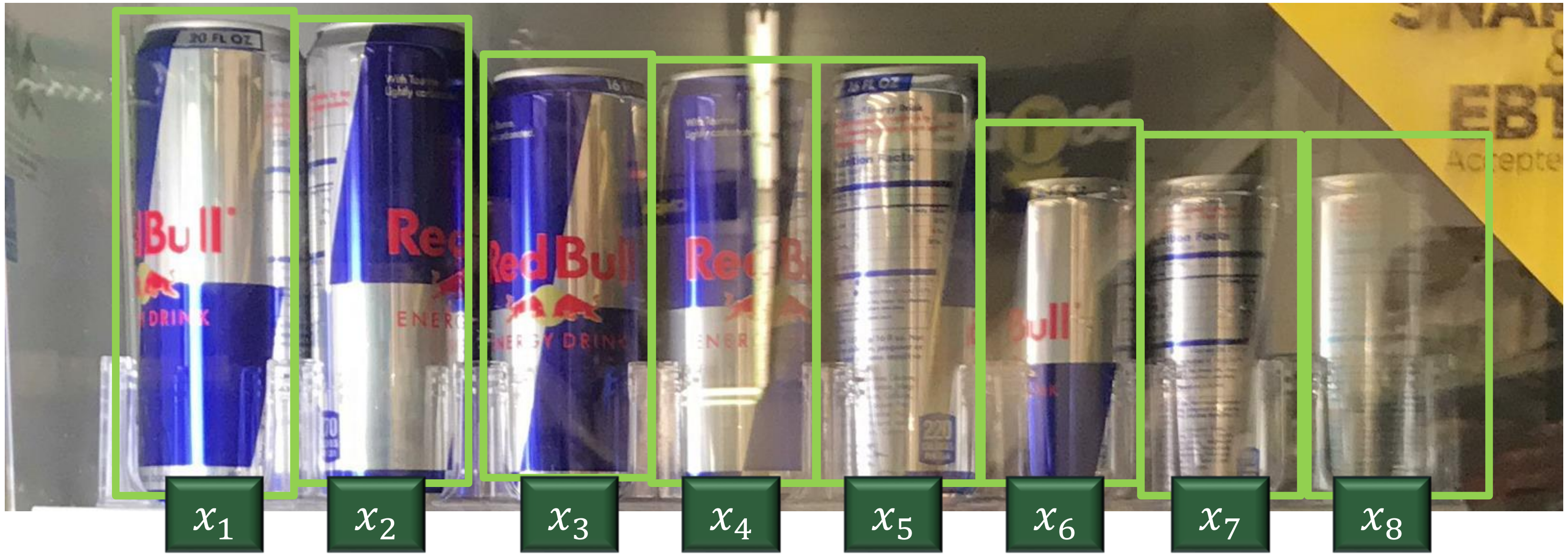
# Classification



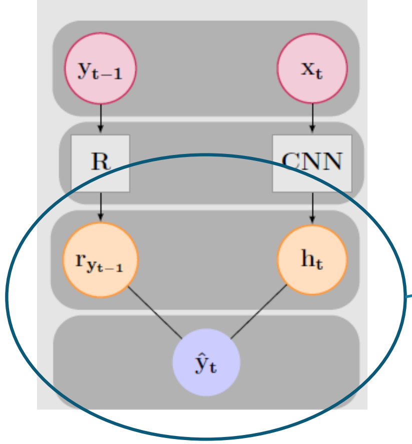

$$P(y_8 | x_1, x_2, x_3, x_4, x_5, x_6, x_7, x_8)$$

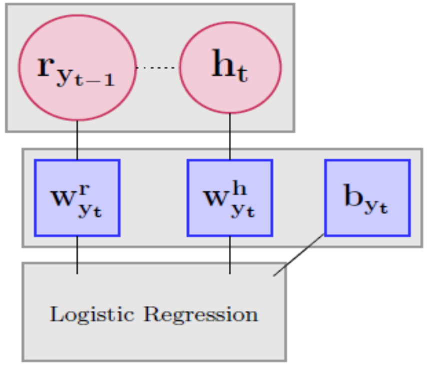

CRF: $P(y_t|y_{t-1}, x_t)$
(One-Hot Vector, Image Matrix)
Input:
$y_{t-1}$
$x_t$
Hidden Layers:
R
CNN
Representations:
$r_{y_{t-1}}$
$h_t$
Classifications:
$\hat{y}_t$

Joint Probability Distribution

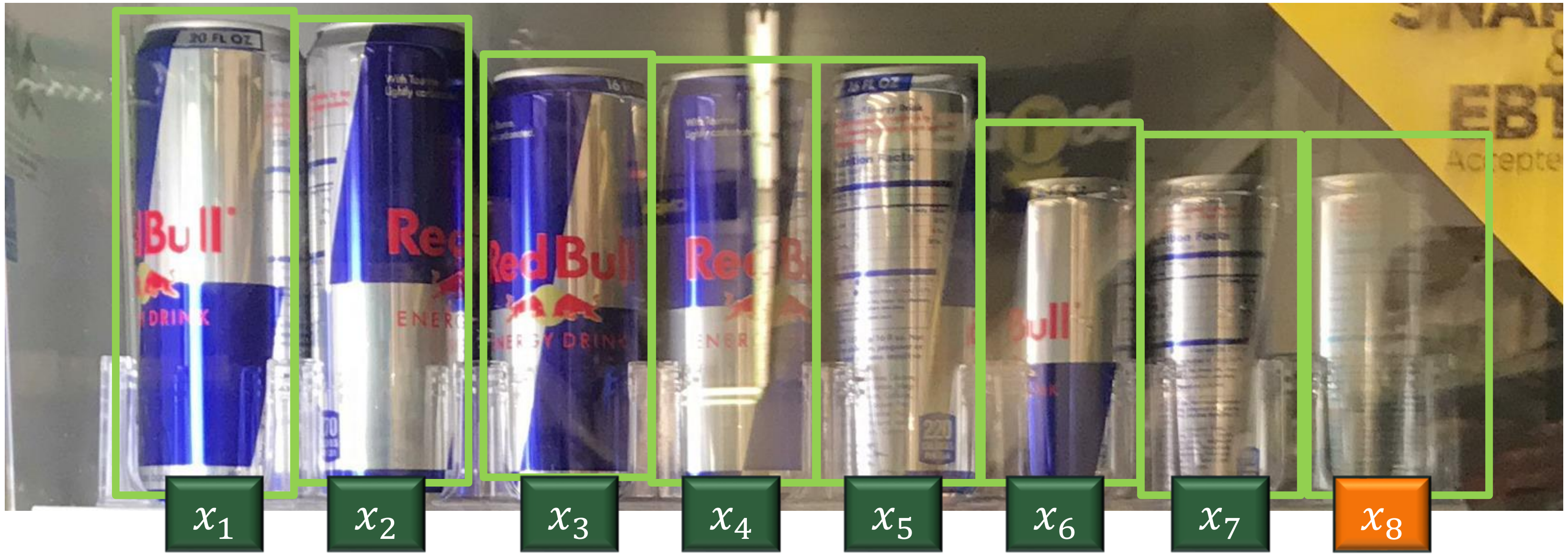



$$P(Y/X) \propto \prod_{t=1}^{n} \varphi(y_t | x_t, y_{t-1}) \propto \prod_{t=1}^{n} e^{[r_{y_{t-1}}^T, h_t^T] W_{y_t} + b_{y_t}}$$



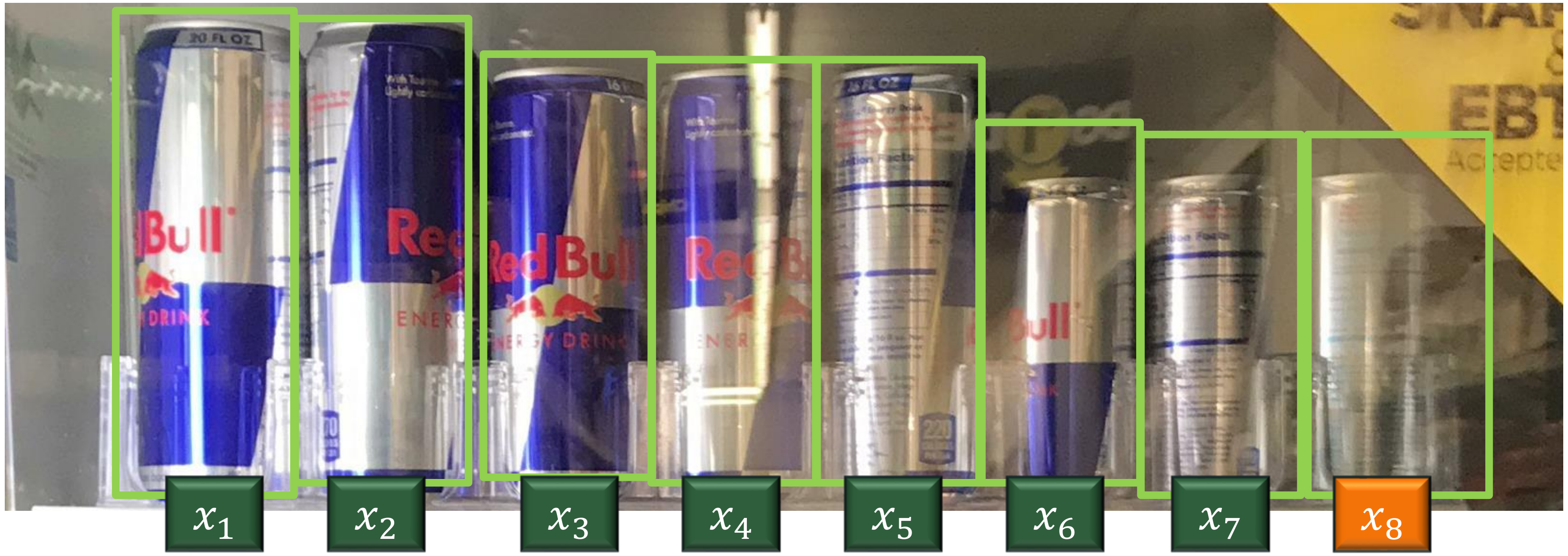
Marginal Distributions




$$P(y_t = i|X) = \sum_{Y/y_t} P(Y/X) =$$

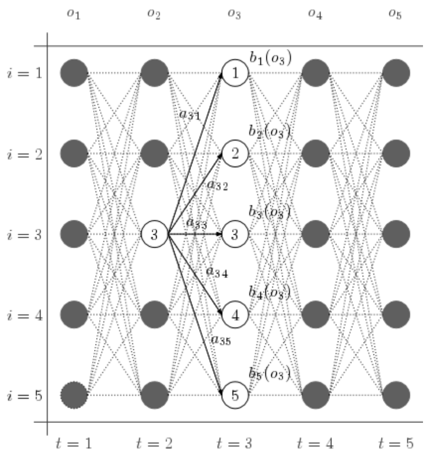
$$\sum_{Y/y_t} P(y_1, y_2, \cdots, y_{t-1}, i, y_{t+1}, \cdots, y_n | x_1, x_2, \cdots, x_n) \propto$$

$$\sum_{Y/y_t} \prod_{\tau=1}^{n} e^{h_\tau^T u_{y_\tau} + b_{y_\tau}} e^{r_{y_{\tau-1}}^T q_{y_\tau}} = \alpha(y_t) \cdot \beta(y_t)$$

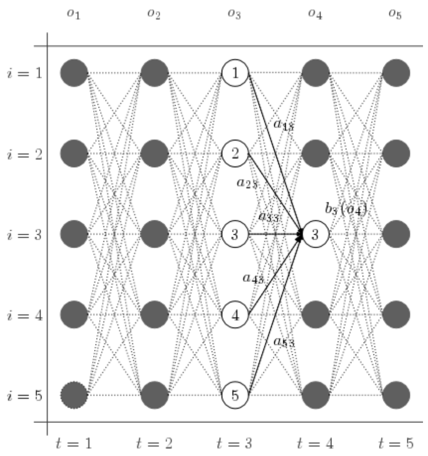



Dynamic programming:
Forward–backward algorithm

$$P(y_t|X) = \frac{1}{Z}\alpha(y_t)\beta(y_t)$$




$\alpha(y_t)$




$\beta(y_t)$

**Classification:** $P(y_i | x_i)$

Spanning Tree Belief-Propagation


Classification: $P(y_i|X)$

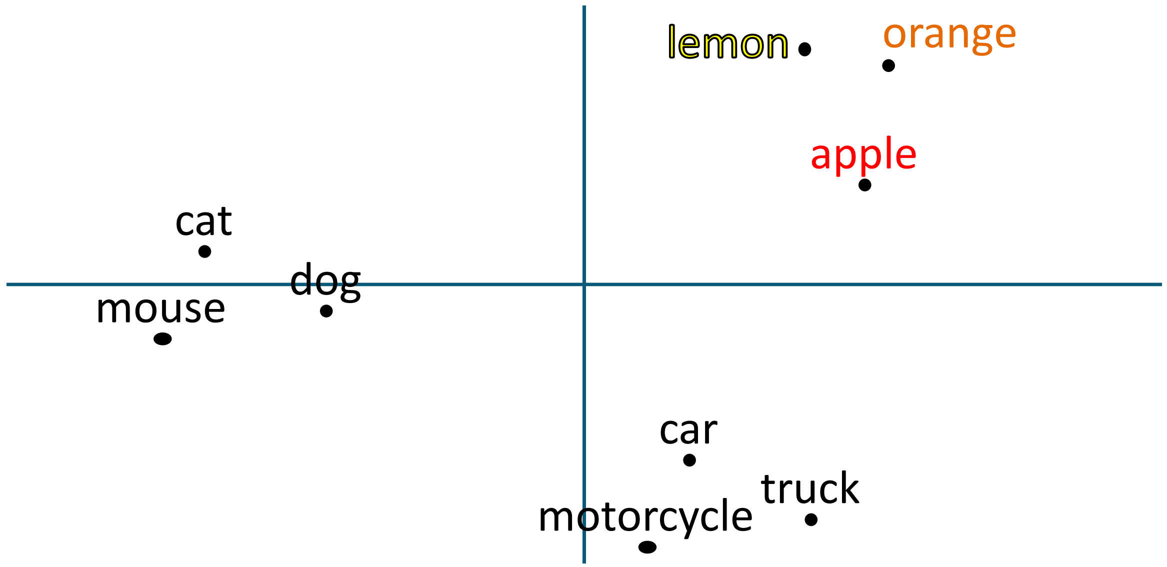



Trax image recognition

# Word Embedding $2^{nd}$ Order Similarity

*Orange juice*
*Orange water*
*Orange tree*
*Orange seed*
*Orange peel*
*Orange pulp*
*Orange orchard*
*Orange milkshake*
*Orange pie*
*Orange soda*
*Orange plantation*
*Orange drink*
*Orange flavored*
*Orange cider*
*Orange vinegar*

*Lemon juice*
*Lemon water*
*Lemon tree*
*Lemon seed*
*Lemon peel*
*Lemon pulp*
*Lemon orchard*
*Lemon milkshake*
*Lemon pie*
*Lemon soda*
*Lemon plantation*
*Lemon drink*
*Lemon flavored*
*Lemon cider*
*Lemon vinegar*






Trax image recognition

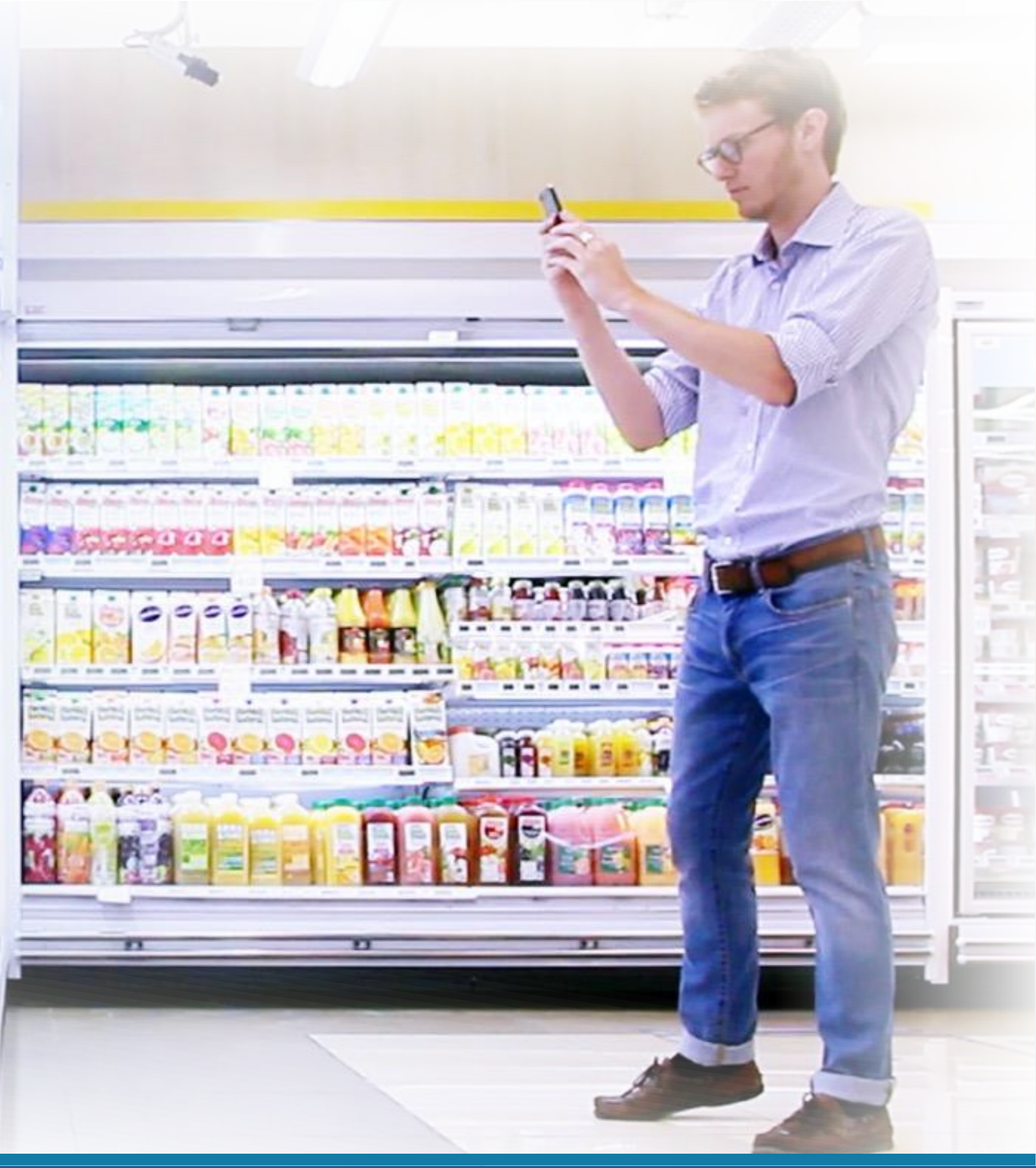
# Product Embedding $2^{nd}$ Order Similarity



Trax image recognition

# AGENDA

- Trax Visual Challenges
- Deep Context Embedding Architecture
- Implementation Details
- The Detection Challenge
- Summary






Trax
image recognition

# Correct Detection

# Missing an object

you shall ~~not~~ pass





Trax image recognition

# Splitting an object

## ......the fireworks on the fourth of July




Trax image recognition

# State of The Art Detectors

“YOLO imposes strong spatial constraints on bounding box predictions”

*Redmon, Joseph, et al. "**You only look once**: Unified, real-time object detection."*

“SSD is very sensitive to the bounding box size”

*Liu, Wei, et al. "**SSD**: Single shot multibox detector."*

“Trying to directly regress to constitutes a difficult learning task”

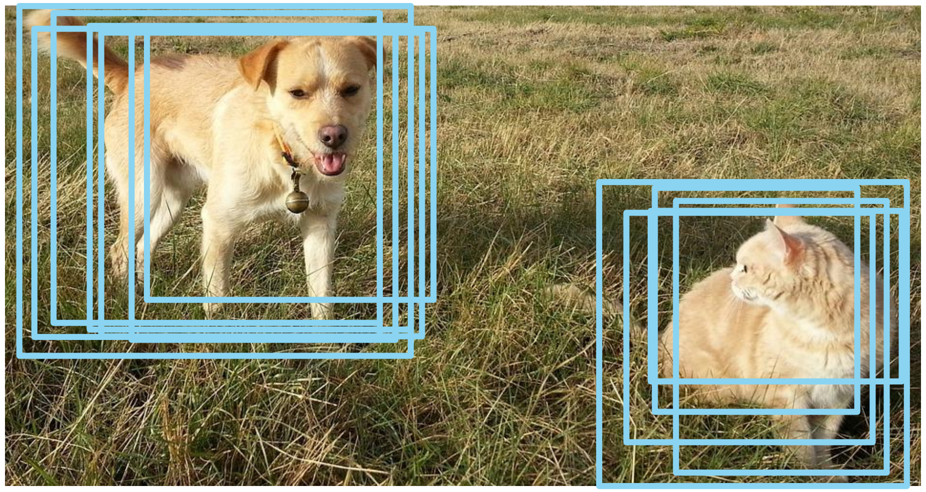
*Gidaris, Spyros, and Nikos Komodakis. "**Locnet**: Improving localization accuracy for object detection."*

“If we pick better priors for the network to start with we can make it easier”

*Redmon, Joseph, and Ali Farhadi. "**YOLO9000**: Better, Faster, Stronger."*

# Duplicate Merger

The “Standard” Case



Our Case

# Are these products  cans?

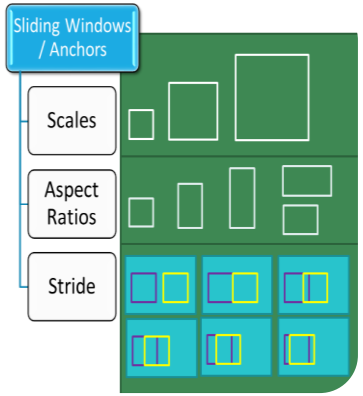
Yes!

Not Really

No!

Really Not

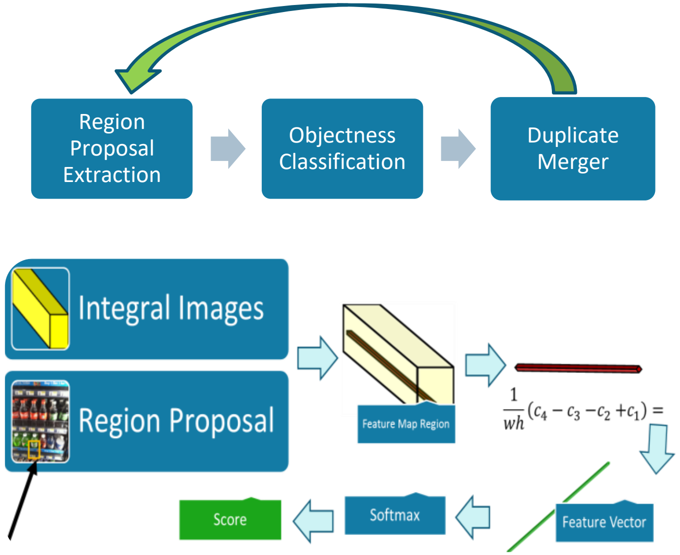
# Detector Innovations - Poster




Region Proposal Network

Objectness Soft Labels

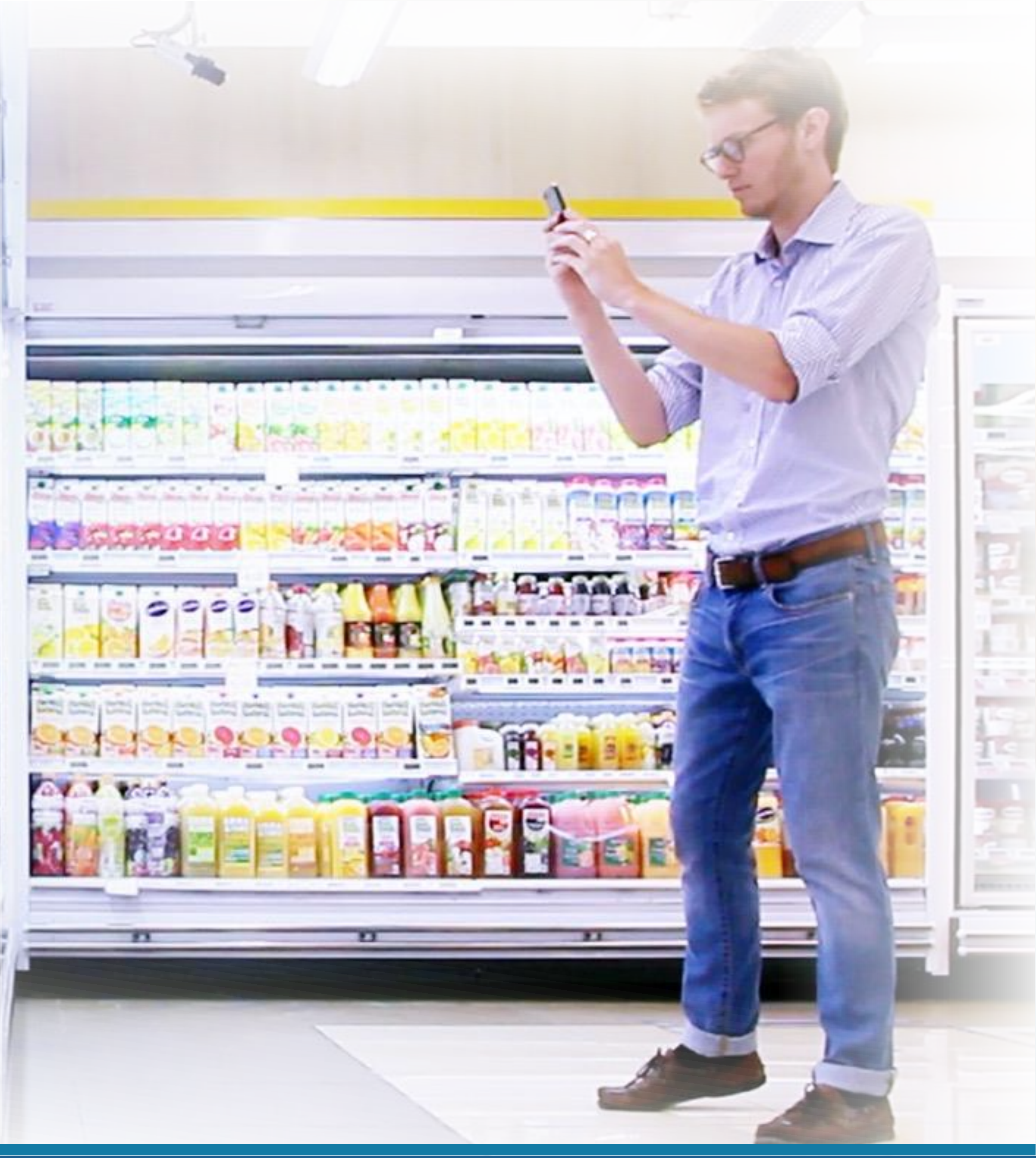
Robust Duplicate Merger




O(1) Region Classification

# *AGENDA*

- Trax Visual Challenges
- Deep Context Embedding Architecture
- Implementation Details
- The Detection Challenge
- Summary



Trax image recognition

# *Take Home Message*

- Fine-grained Classification is Challenging
- New Context Embedding CNN Architecture
- NLP Inspired
- Detection challenge

Trax
image recognition

Thank You