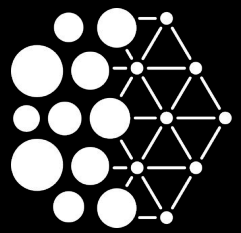
Institut
québécois
d'intelligence
artificielle



Mila

Humanware project
Detection of text in natural scenes

Margaux Luck, PhD
Jeremy Pinto

HumanWare

Over 25 years of service to people with vision loss

Recognized for its capacity to innovate
(Digital reading systems, talking book players, handheld talking GPS)

A worldwide network of distributors
(North America, Europe & Australasia)

150 employees making independence possible
(15% of our employees are also our clients)

#1 worldwide in blindness service

Products supporting over 25 languages
(including the Braille script)

Products available in more than 45 countries

Drummondville, Canada
Head Office
Operations Canada & USA

Rushden, Angleterre
Operations Europe,
Middle East & Africa

Longueuil, Canada
Marketing & R&D

Sydney, Australia
Operations Australasia

Humanware project

- **Task:** Detection of text in natural scenes, with its location and the ability to guide the user to get the full text (alignment) so that the text is interpretable by a recognition engine.

- **Use case:** Detection of house numbers, text on a bus stop sign, or text on a storefront (name, product details, opening hours, ...)

- **Constraints:** Execution time, online vs. offline, memory usage (in the case of a mobile application), etc.

Focus on door number detection

668

- Help blind persons find their way around
- Make sure that’s the right house
- (Long term) Facilitate micronavigation

Data

The Street View House Numbers (SVHN) Dataset:

- Open-source
- ~200k street numbers
- bounding boxes and class labels for individual digits, giving about 600k digits total

Limitation:

- zoom on the numbers, lack of background
- no negative examples (i.e. no images without numbers)

[Yuval Netzer, Tao Wang, Adam Coates, Alessandro Bissacco, Bo Wu, Andrew Y. Ng Reading Digits in Natural Images with Unsupervised Feature Learning NIPS Workshop on Deep Learning and Unsupervised Feature Learning 2011.]

Ongoing data collection

In house dataset

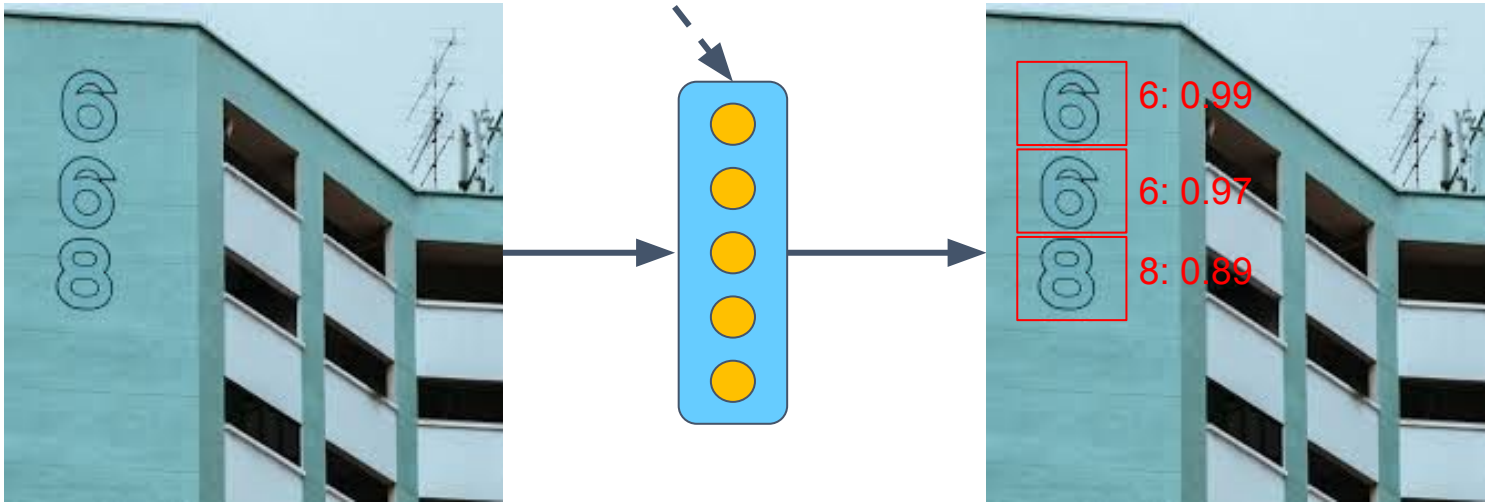
- Aim to be open-source
- Video of street views (case studies defined with the users)
- No ground truth bounding boxes available
- Centroid of the door numbers and sequence label
- The collection started
- You can help!

Object detection

Input image

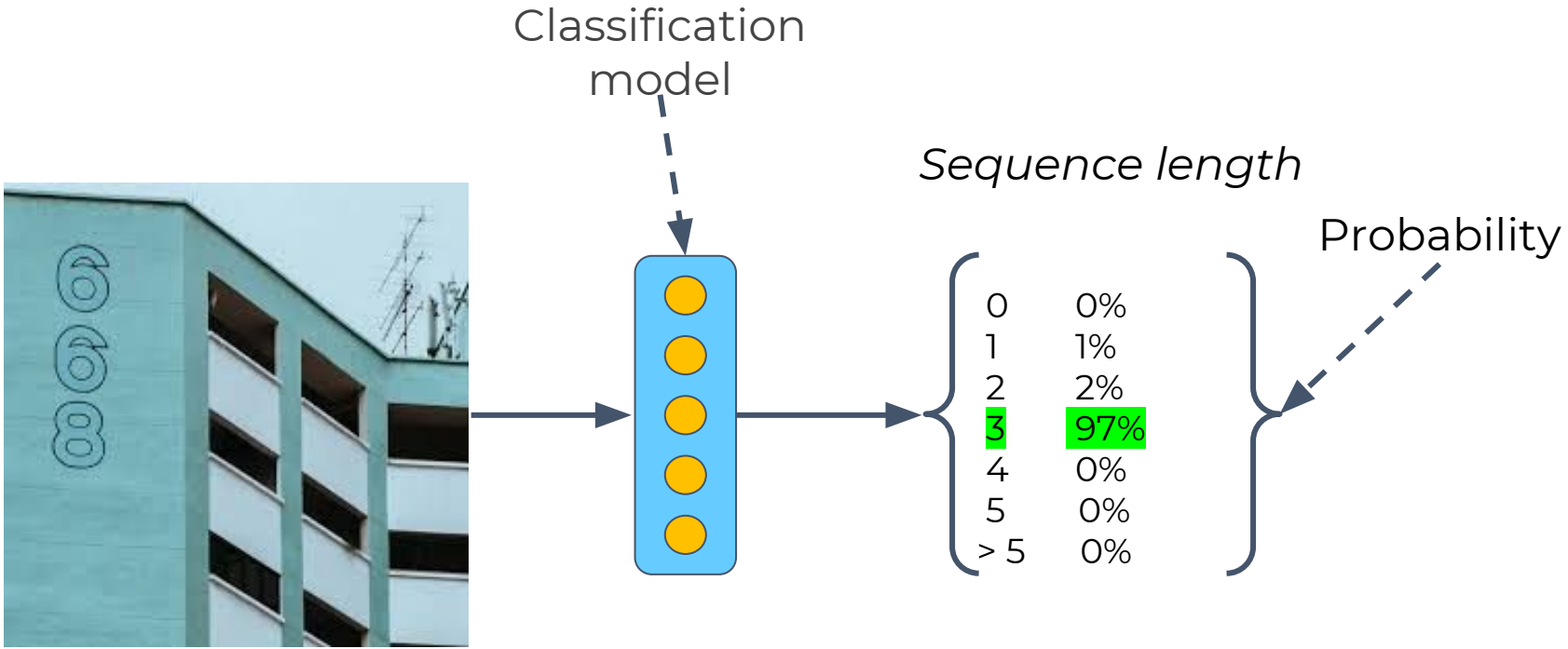
Object detection model

Bounding boxes' coordinates
+ Digit presence probability
+ Probability of recognized digits

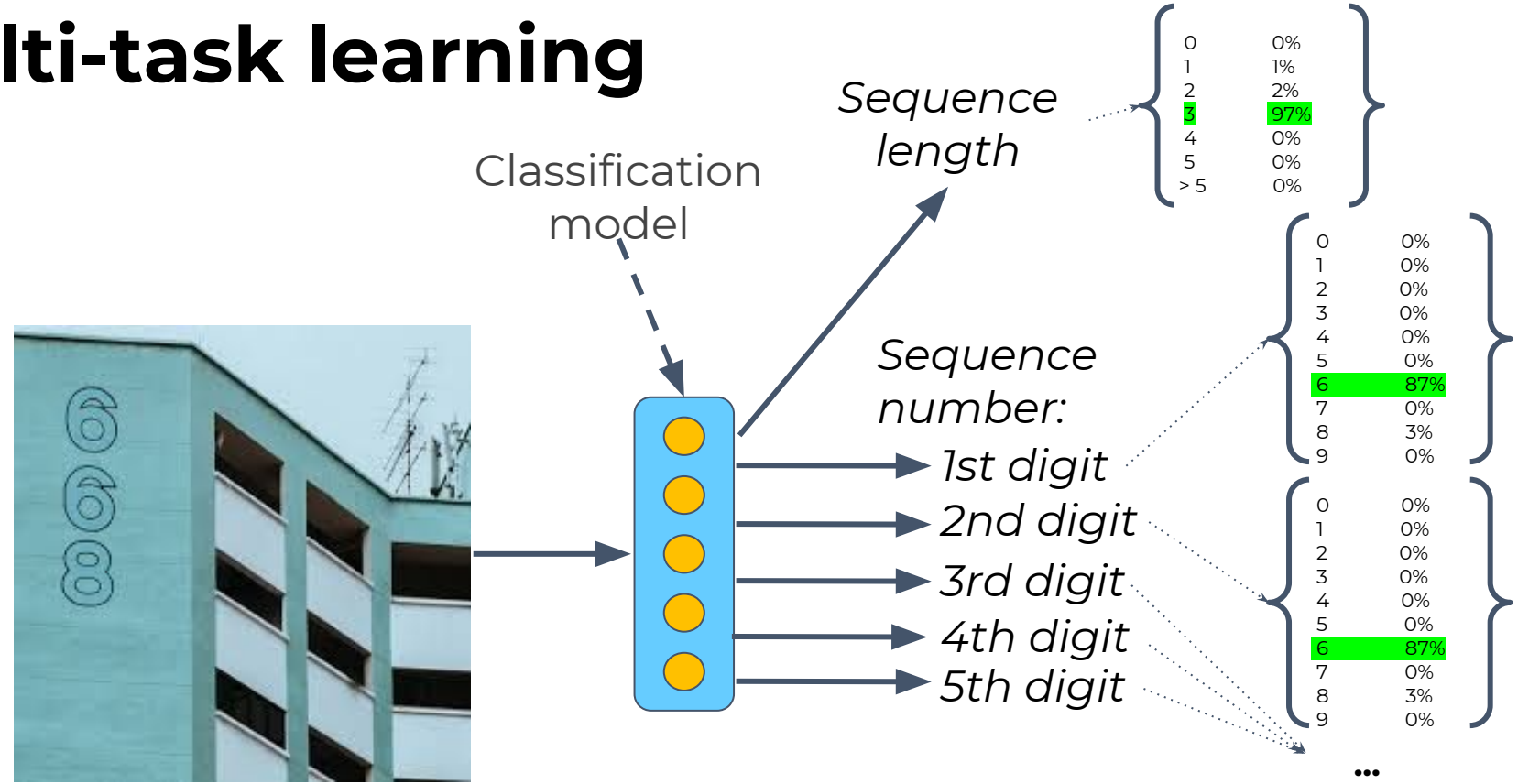


Mila

Classification



Multi-task learning



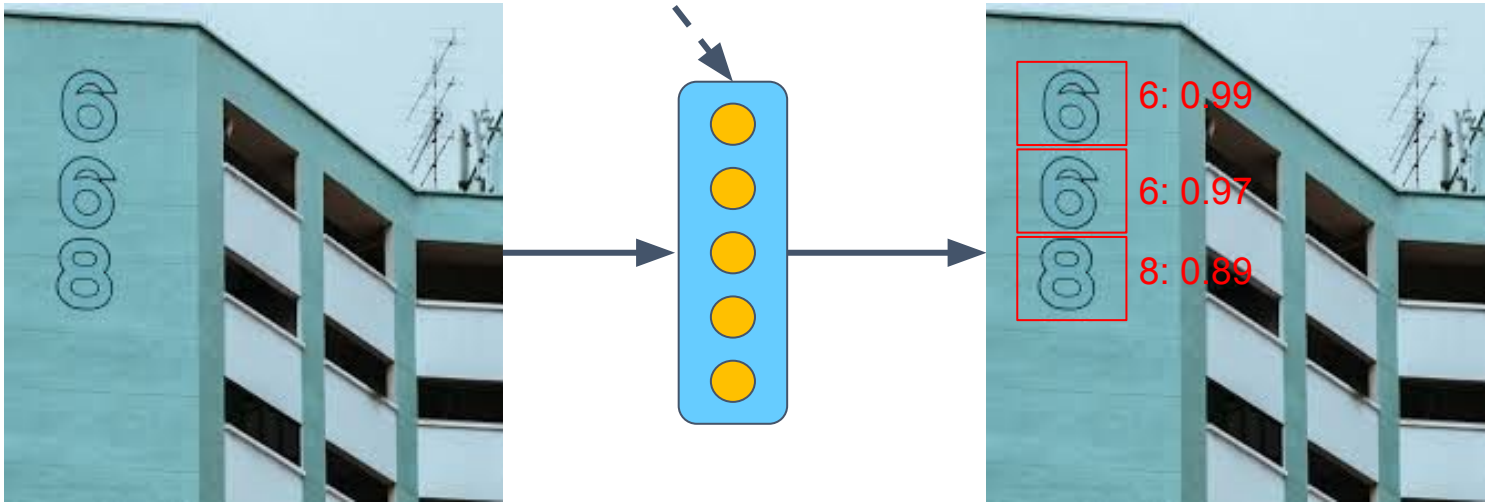
Mila

Object detection (multi-task learning)

Input image

Object detection model

Bounding boxes' coordinates
+ Digit presence probability
+ Probability of recognized digits



Mila

Official evaluation metrics

- Sequence length accuracy as initial metric

That will be replaced with:

- Sequence transcription accuracy
  - # correct sequences / total # of sequences

Informative evaluation metrics

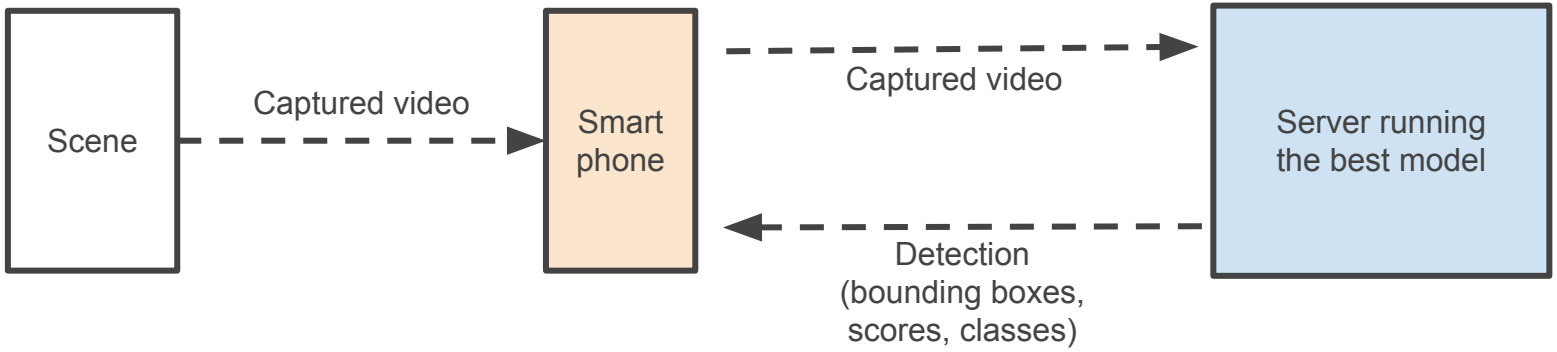
- Sequence length accuracy
- Sequence transcription accuracy
- Digit-level accuracy
- Coverage at 98% sequence transcription accuracy
- etc.

\+ *For object detection models (that include bounding boxes)*

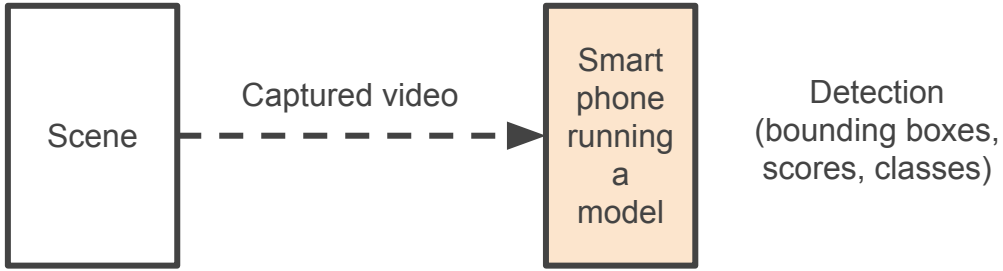
- Sequence intersection over union (IoU)
  - IoU = (area of overlap) / (area of union)
- Digit-level IoU
- etc.

Demo

Door number detection: server solution



Door number detection: embedded solution



Mila