Institut québécois d'intelligence artificielle

Mila

Door number detection project
Detection of text in natural scenes

Margaux Luck, PhD
Jeremy Pinto

Door number detection 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.

Mila

Focus on door number detection

668

- Help blind persons find their way around

- Make sure that's the right house

- (Long term) Facilitate micronavigation

Mila

Data (block 1 & 2)

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.]

Mila

Data (block 3)

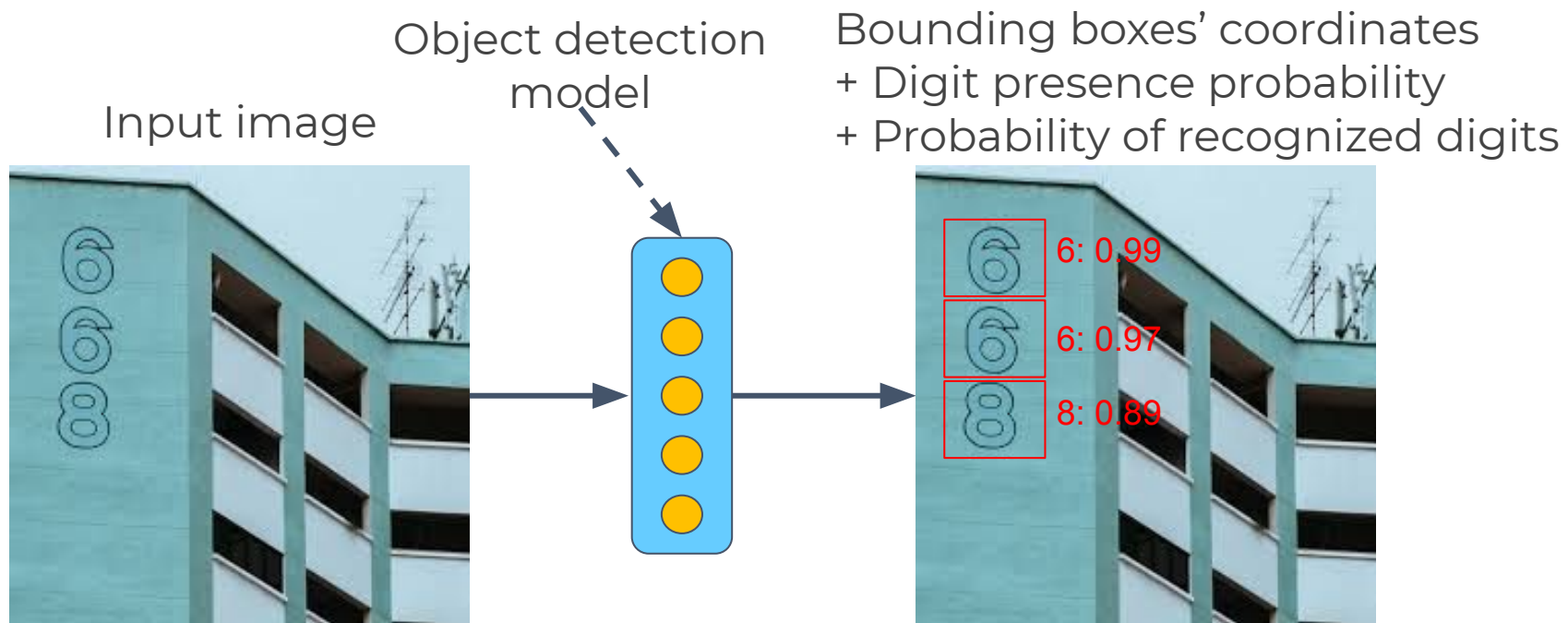*Synthetic dataset (kindly provided by ElementAI):*
- Private for now
- 6k street numbers (5k train + 1k valid)
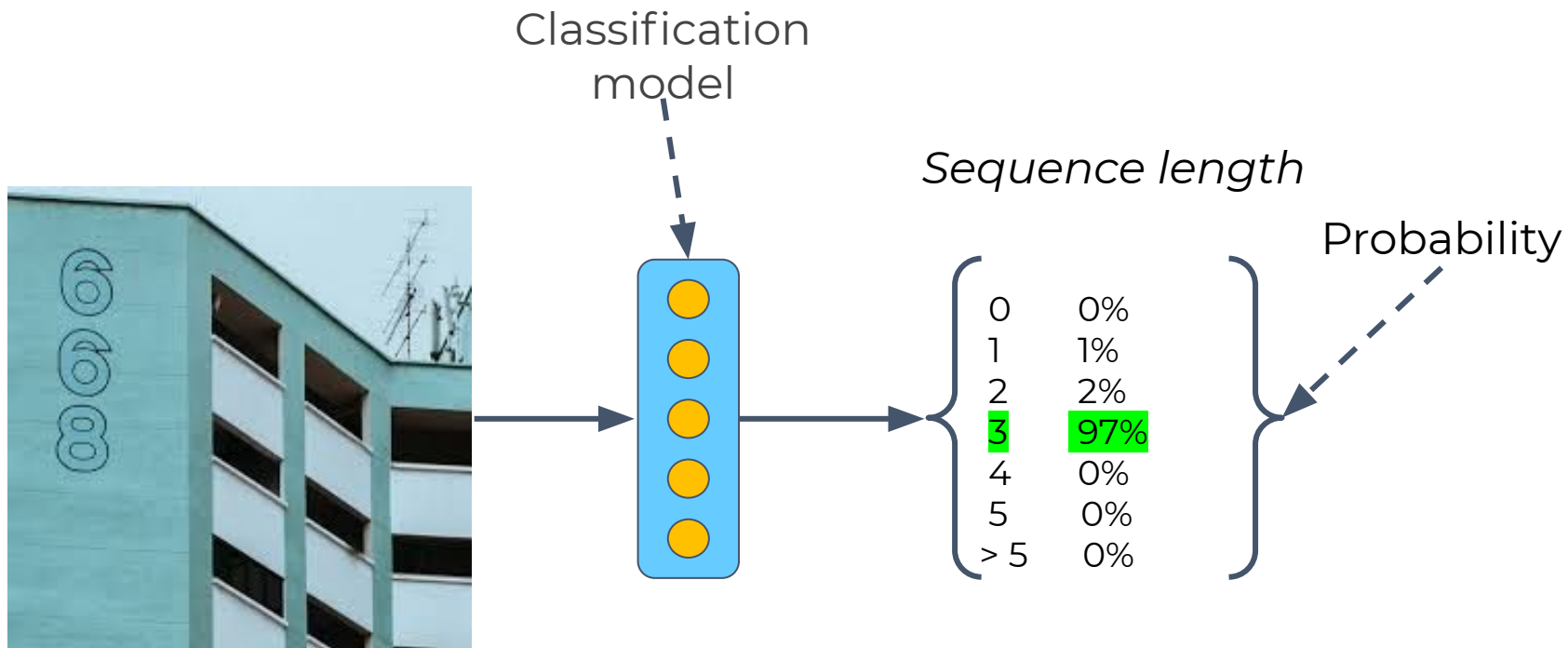- bounding boxes, segmentation masks and class labels for sequence of digits

Limitation:
- Not really realistic
- Not enough diversity

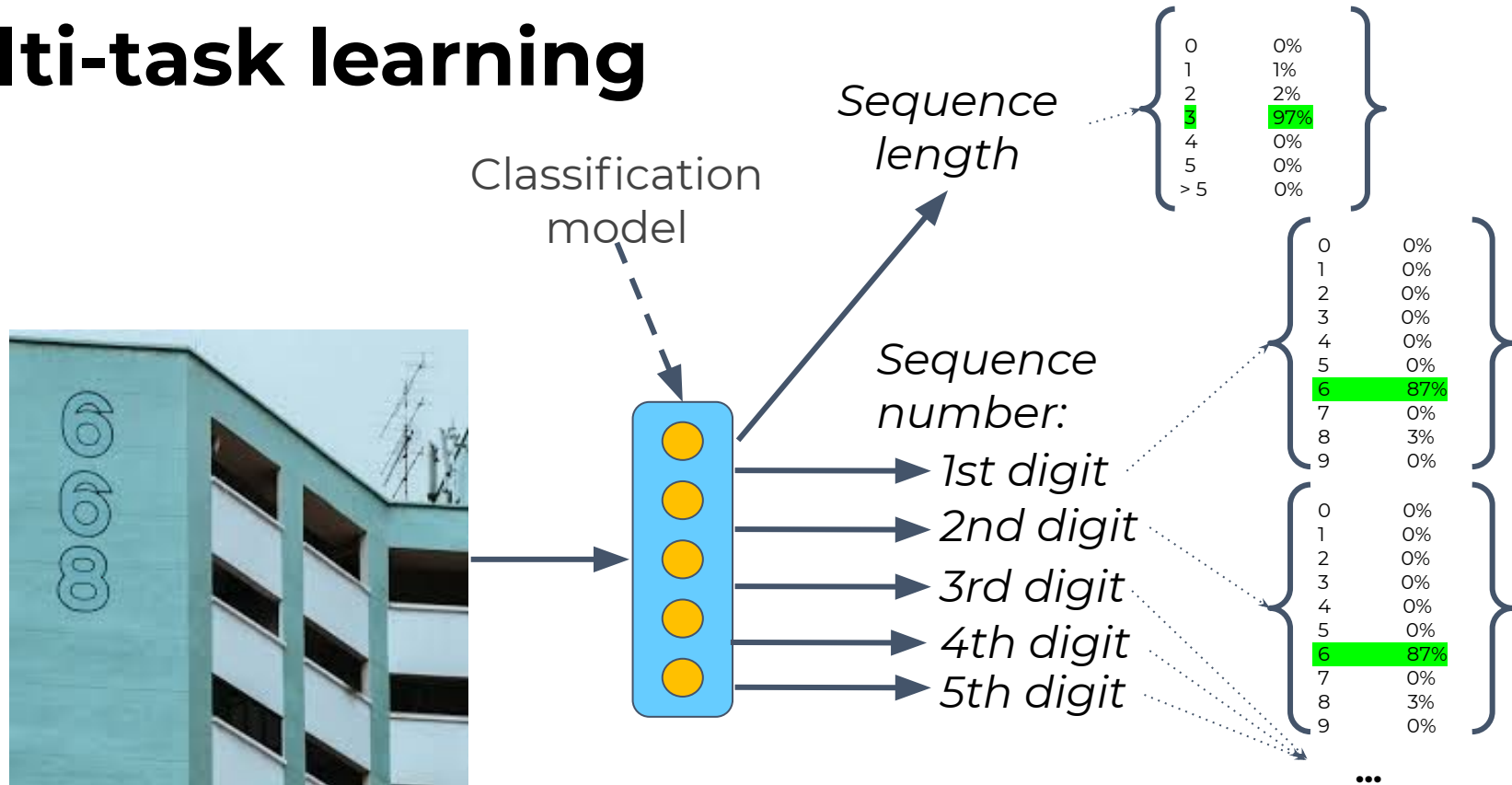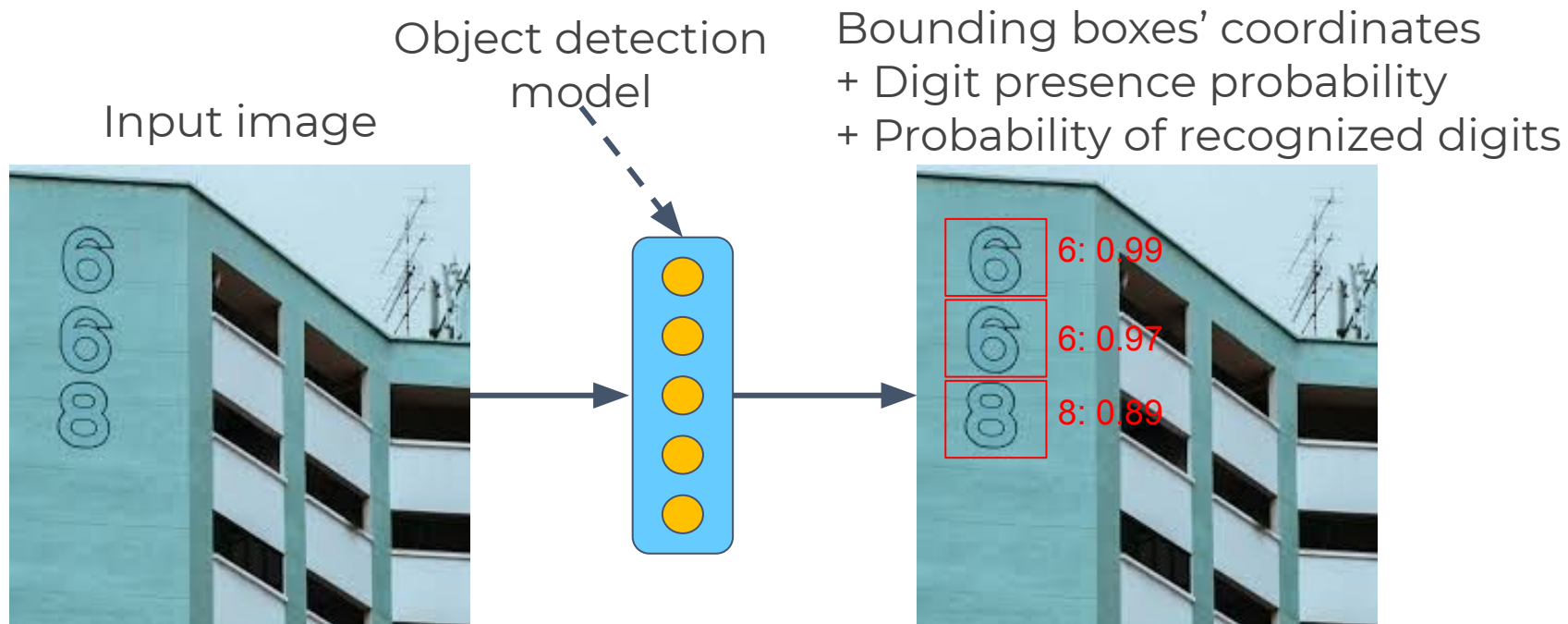... But different angles and more varied viewpoints than SVHN

Object detection

Input image

Object detection model

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

6: 0.99

6: 0.97

8: 0.89

Classification

Classification model

Sequence length

Probability

0	0%
1	1%
2	2%
3	97%
4	0%
5	0%
> 5	0%

Mila

Multi-task learning

Object detection (multi-task learning)

Object detection model

Input image

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

6: 0.99

6: 0.97

8: 0.89

Official evaluation metrics

- Sequence length accuracy as initial metric

That will be replaced with:
- Sequence transcription accuracy
  - # correct sequences / total # of sequences

Mila

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
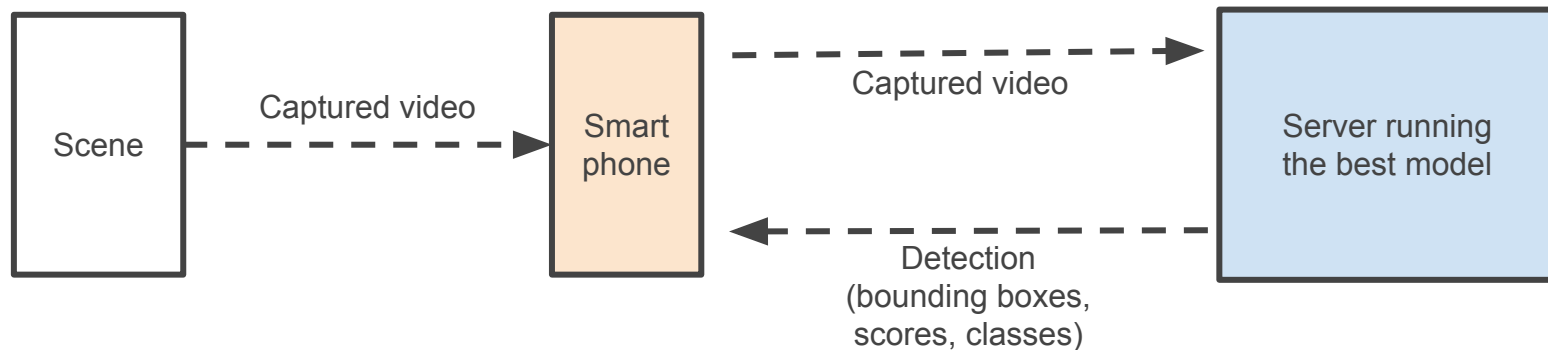
Door number detection: embedded solution