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.

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

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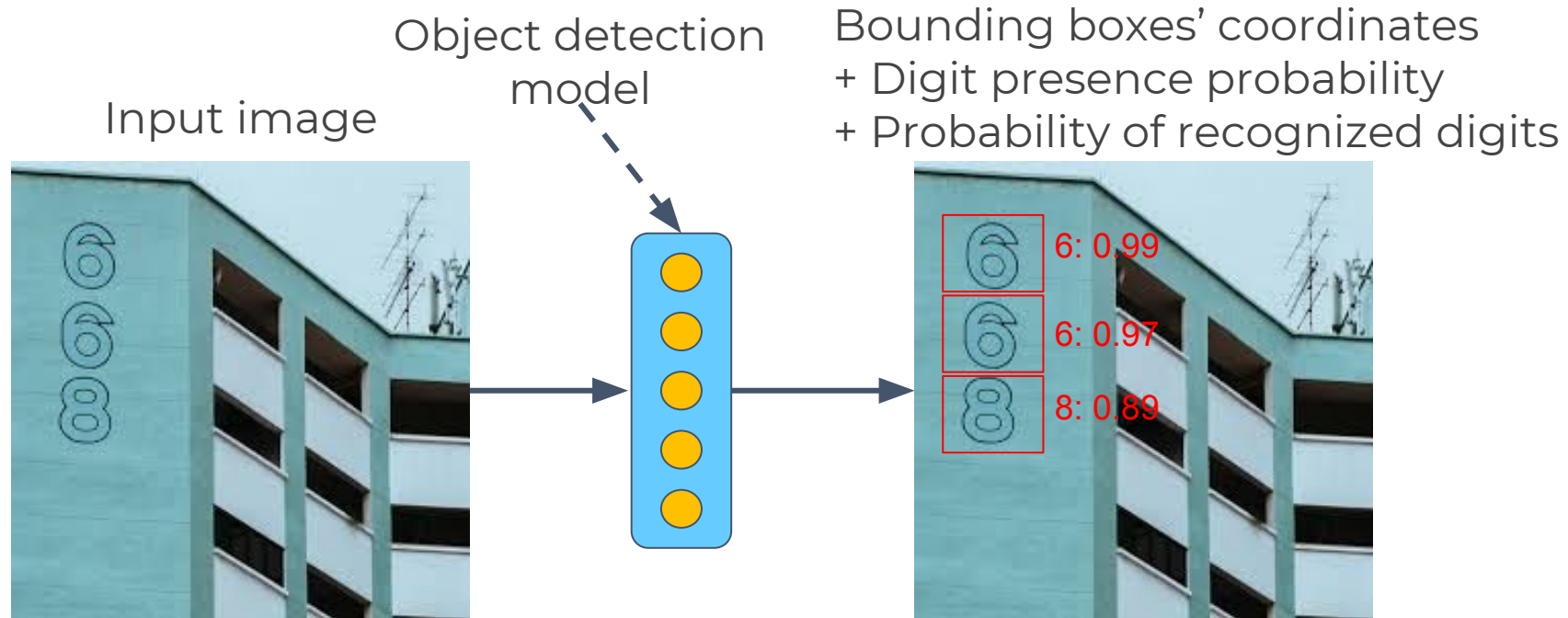
Ongoing data collection
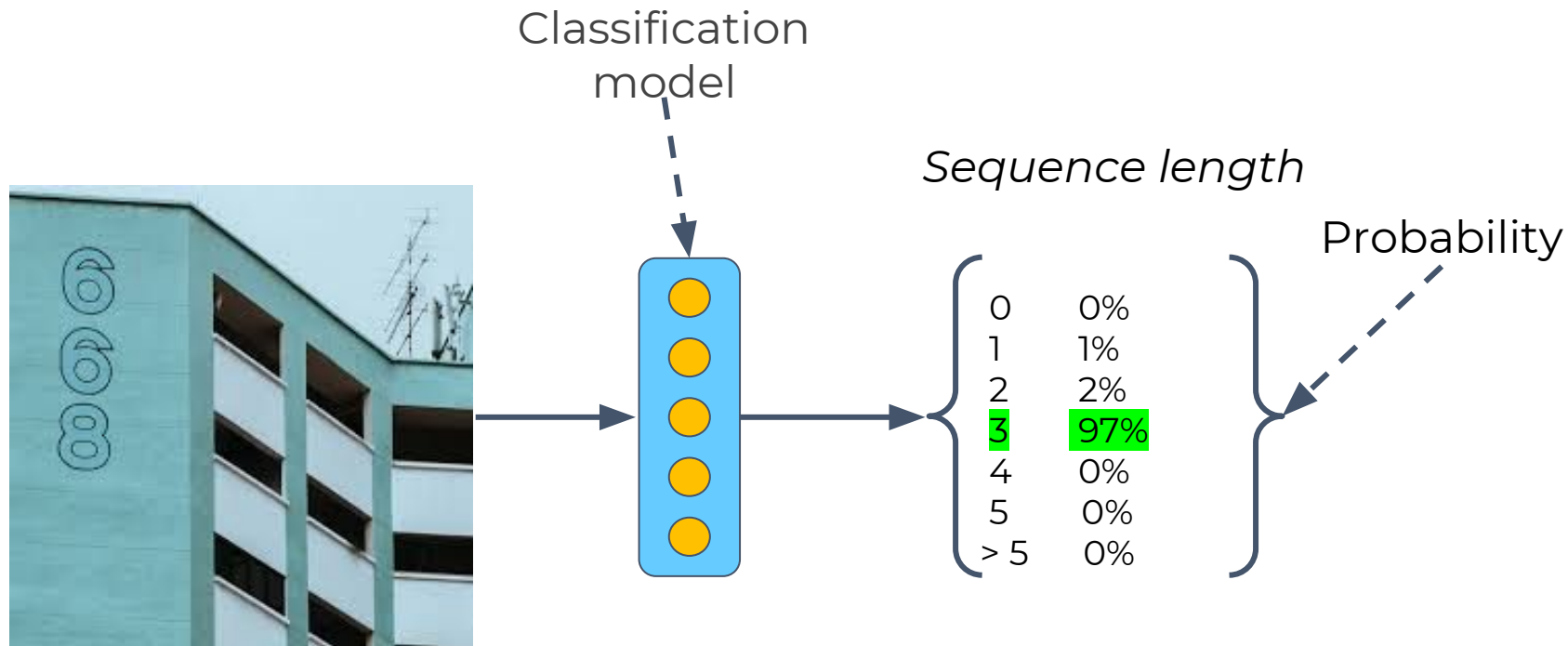
In house dataset
- Aim to be open-source
- Video of street views (case studies defined with the users)
- No ground truth bounding boxes available
- Centroid of the door numbers and sequence label
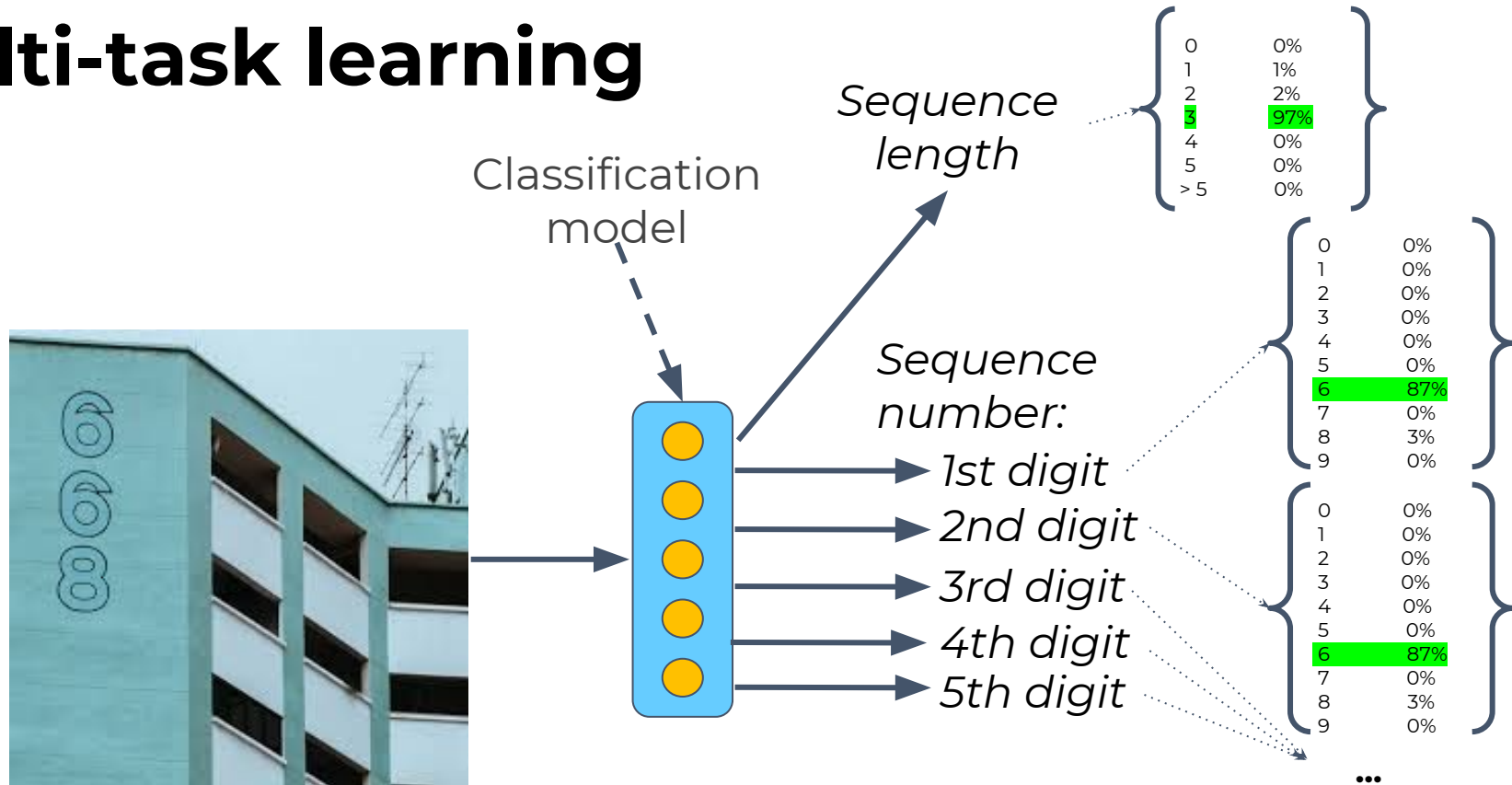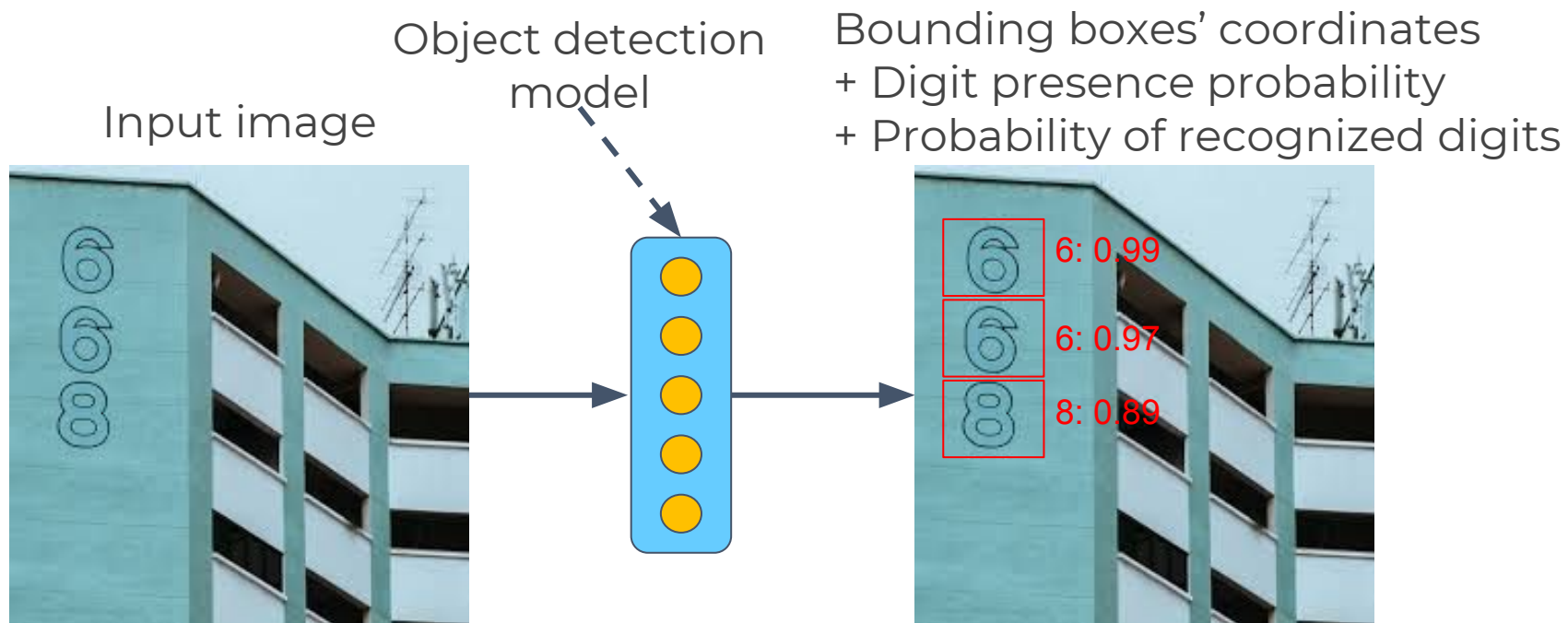- The collection started
- You can help!

Object detection

Input image

Object detection model

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

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%

Multi-task learning

Classification model

Sequence length

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

Sequence number:

1st digit

2nd digit

3rd digit

4th digit

5th digit

0	0%
1	0%
2	0%
3	0%
4	0%
5	0%
6	87%
7	0%
8	3%
9	0%

0	0%
1	0%
2	0%
3	0%
4	0%
5	0%
6	87%
7	0%
8	3%
9	0%

...

Mila

Object detection (multi-task learning)

Object detection model

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

Input image

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
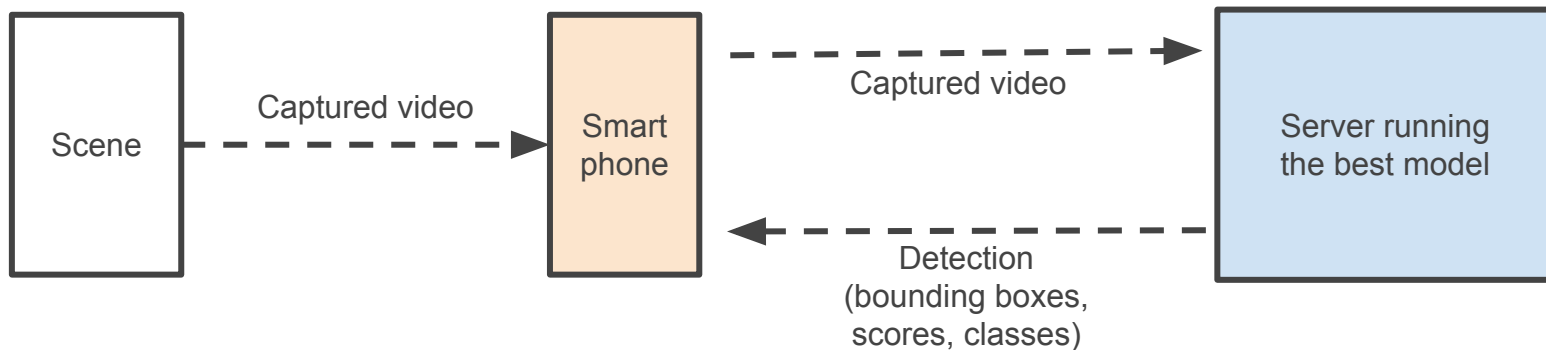
Door number detection: embedded solution

Mila
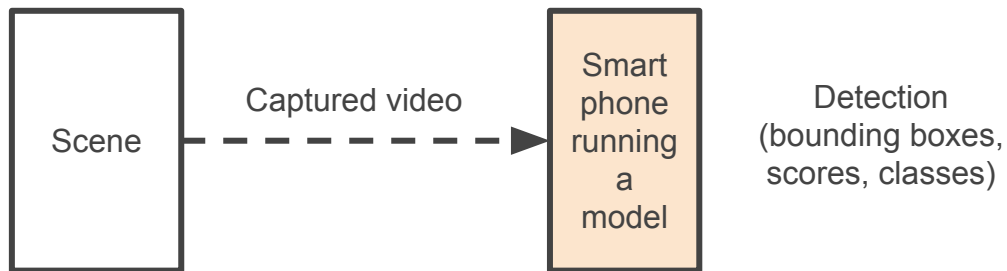
Block 1 instructions / expected timeline

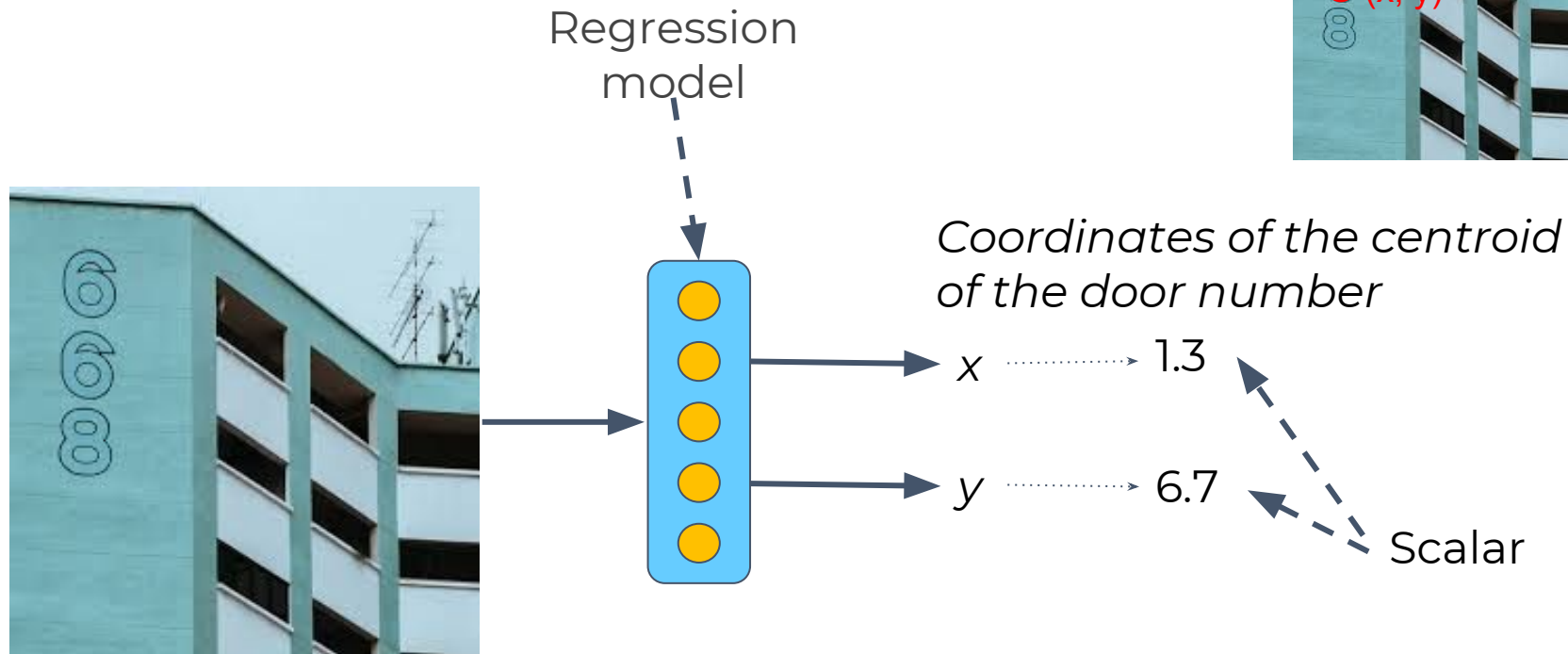|	2019/01/14 week	2019/01/21 week	2019/01/28 week	2019/02/04 week
Tasks / Homework	• Literature review of available datasets • Read the [Goodfellow et al. 2013](#) article	• Code the data loader for SVHN as in Goodfellow et al. 2013 • Implement data visualization functions	• Train a CNN for the classification of the sequence length	• Write a short report summarizing the work, and results • (Peer-) Review of other teams' code
Objectives / Deliverables	• Justify the choice of using the SVHN dataset and to collect an in-house dataset	• SVHN data loader	• CNN for the classification of the sequence length	• Produce documented code and report summarizing the experimental work • Provide model for blind test set evaluation • Complete the peer code review

Mila

Block 2 instructions / expected timeline

	2019/02/11 week	2019/02/18 week	2019/02/25 week	2019/03/11 week
Tasks / Homework	• Read and understand the documentation (code) and reports	• Implementation of loss to classify the different digits of the image (Multi-Task learning)	• implementation + evaluation of different models (AlexNet, VGG, SqueezeNet,...,) / inference speed, metrics • Search for hyper-parameters, use of TensorboardX, management of various models, comparative performance grids	• Search for hyper-parameters, use of TensorboardX, management of various models, comparative performance grids • Write a short report summarizing the work, and results • (Peer-) Review of other teams' code
Objectives / Deliverables	• Understand the documentation (code) and reports	• Multi-task learning model for the sequence digits recognition	• Find the best set of hyper-parameters and visualize the models training, ...	• Produce documented code and report summarizing the experimental work • Provide model for blind test set evaluation • Complete the peer code review

Mila

Block 3 instructions / expected timeline

	2019/03/18 week	2019/03/25 week	2019/04/01 week	2019/04/08 week
Tasks / Homework	• Read and understand the documentation (code) and reports	• Add the object detection part to the classification model (+ number extraction) • Application and finetuning (i.e. transfer learning) of the model on the Humanware dataset - they are given the dataloader	• Finetuning of the models and obtaining the results for the final presentation	• Preparation of the final presentation of 15 min + 5 min of questions which describes the evolution of the project through the 3 blocks • (Peer-) Review of other teams' code
Objectives / Deliverables	• Understand the documentation (code) and reports	• Object detection model of the door digit	• Finetune object detection model	• Produce documented code and presentation summarizing the experimental work of the 3 blocks • Provide model for blind test set evaluation • Complete the peer code review

Mila

Multi-task learning

Regression model

Coordinates of the centroid of the door number

x ⋯⋯⋯⋯→ 1.3

y ⋯⋯⋯⋯→ 6.7

Scalar

(x, y)

Multi-task learning

Regression / classification model

x
y

Coordinates of the centroid of the door number

Sequence length

1st digit
2nd digit
3rd digit
4th digit
5th digit

Sequence number

Mila