

WHY YOU NEED TO BUILD YOUR OWN DATASET

*and other tips to solve companies
image recognition problems*

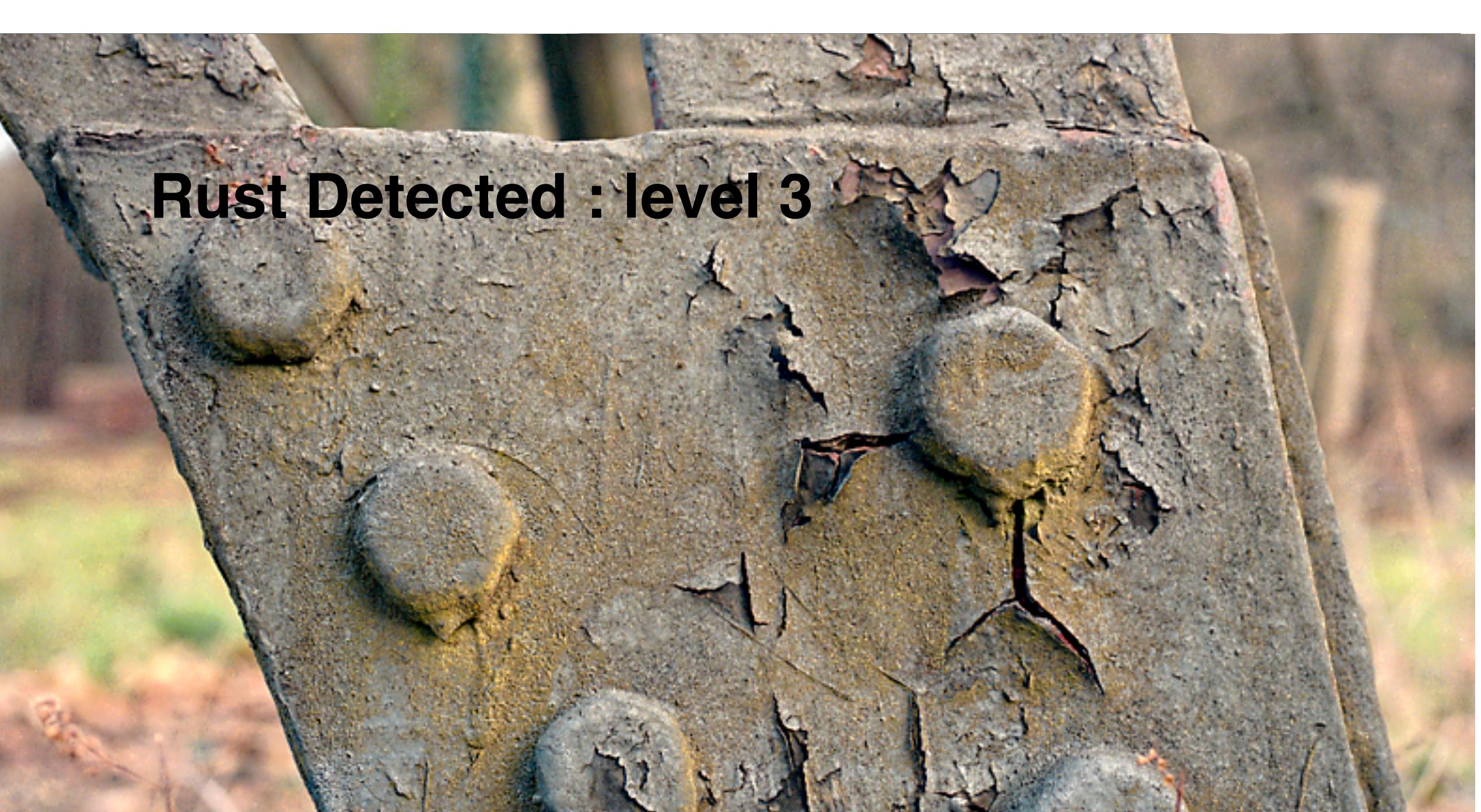
Augustin Marty, CEO & Co-Founder



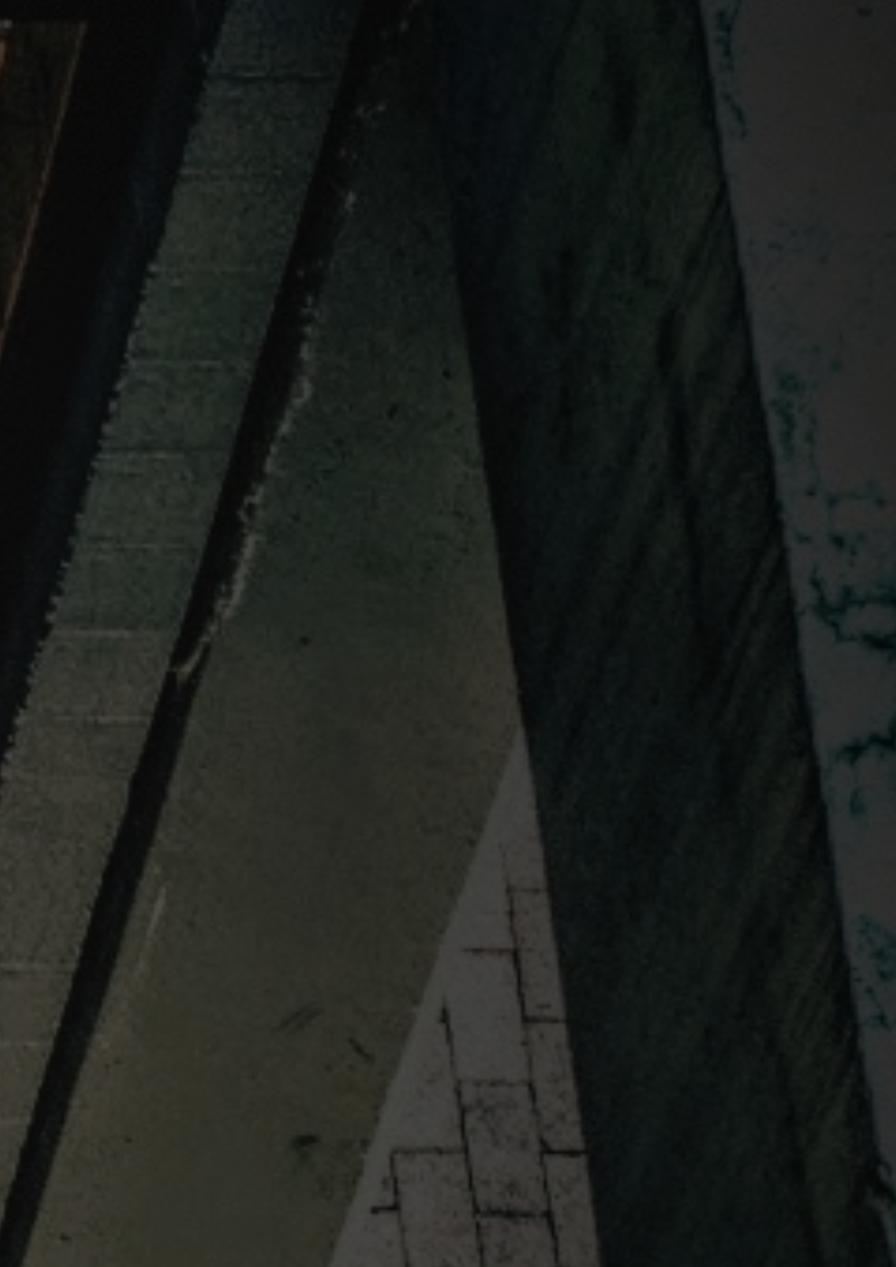
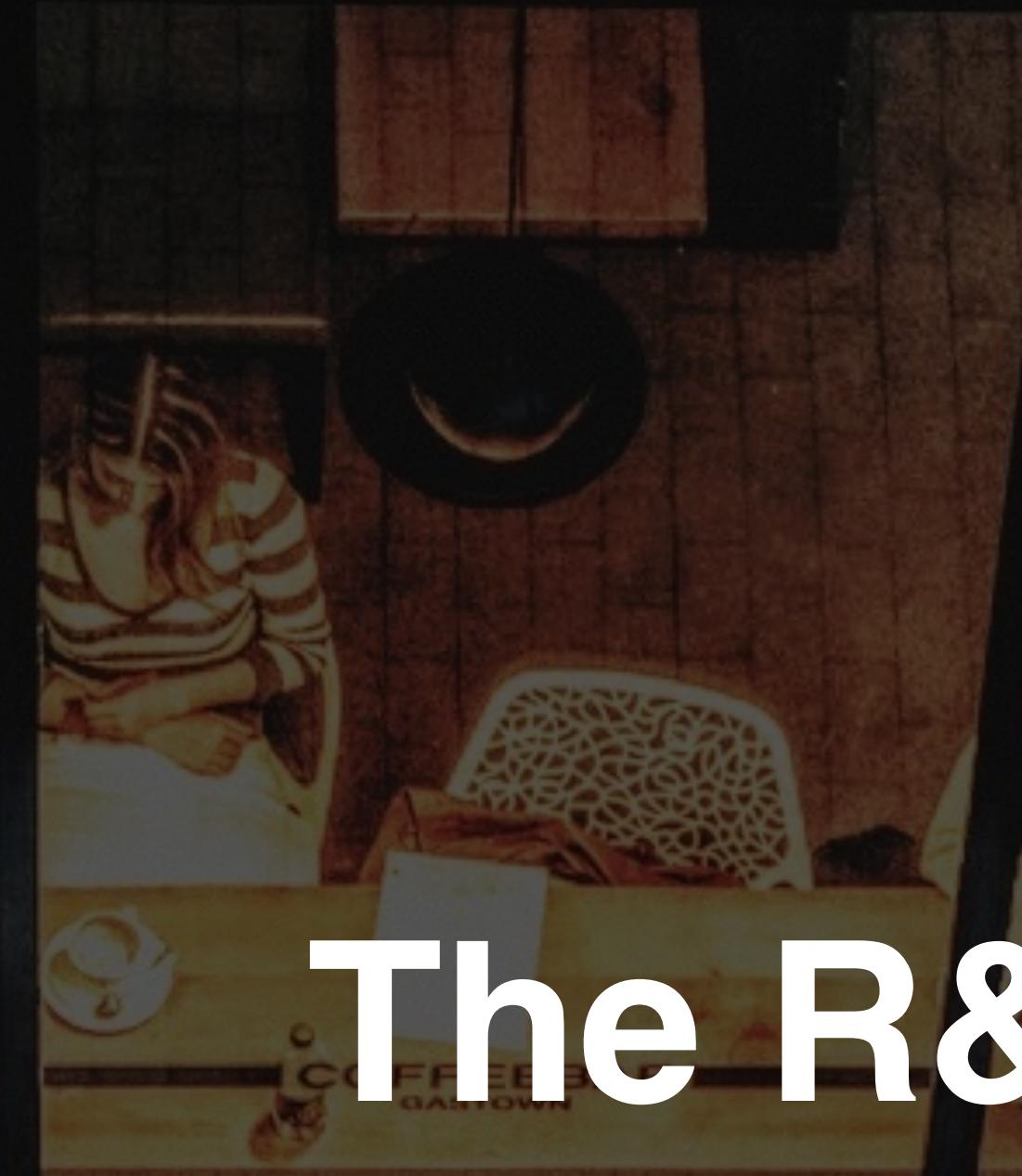
PRODUCT



We develop a B2B software used by companies to **build, train** and **deploy** image recognition systems.



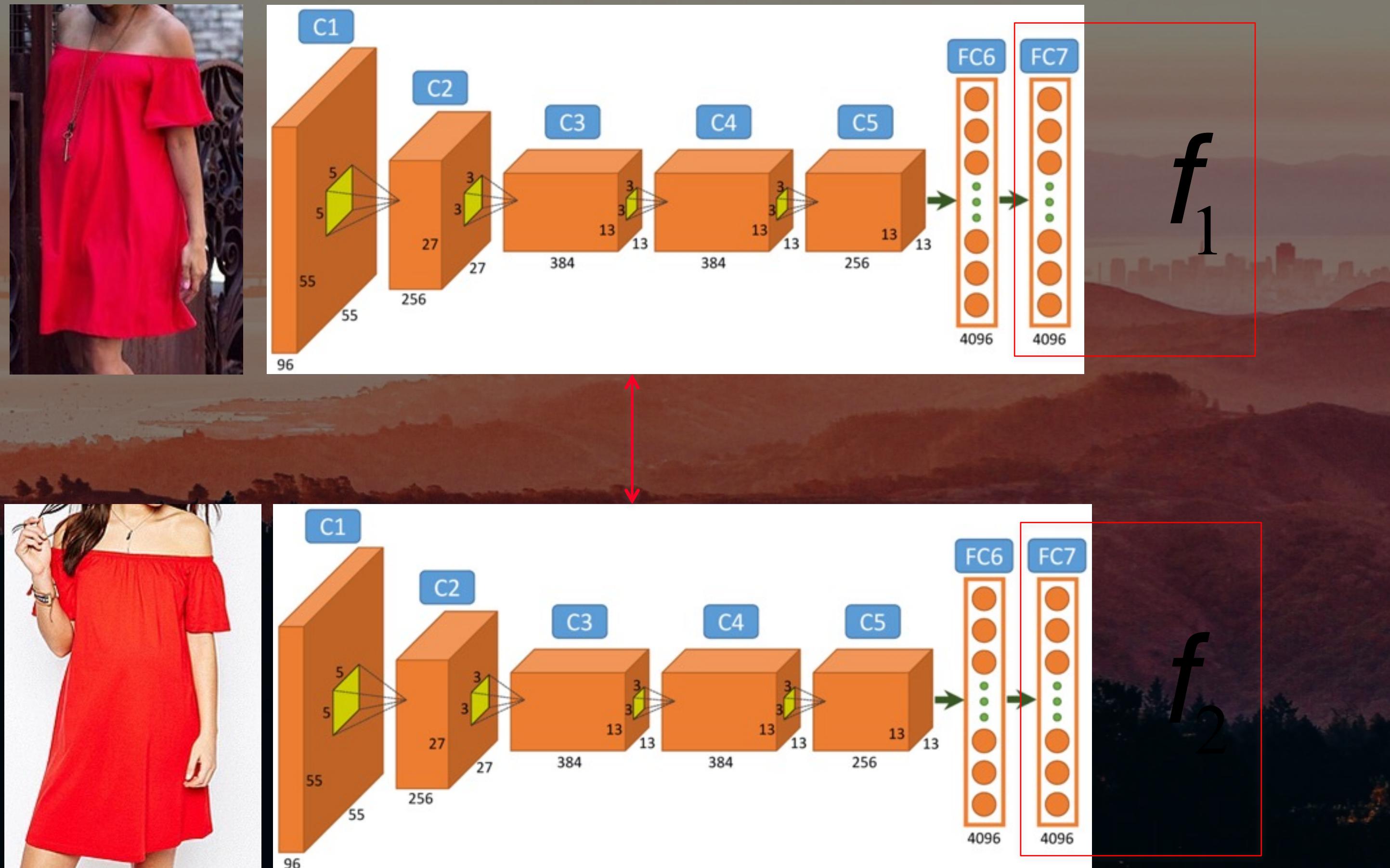
The R&D era



A perfect match database



Siamese Networks



For similar examples, we minimize:

$$L(x_1, x_2) = \|f_1 - f_2\|^2$$



Takeaway #1

*Very hard to compete with state-of-the-art
research from GAFA*



DEEP LEARNING AND COMPUTER VISION LOG BOOK

2017

Arthur Wilcke

Introduction

4

1 Vision Datasets

5

1.1 Detection Datasets	5
1.1.1 Fashion	5
1.1.2 Furniture	5
1.1.3 Weapons	5
1.2 Classification Datasets	5
1.2.1 Compass Main Meals	5
1.2.2 Sacs	6
1.2.3 Condenast	6
1.2.4 Allegro	6
1.2.5 Eyes	6
1.2.6 Vinci	6

2 Data-related modifications

7

2.1 Image pre-processing	7
2.2 Annotation quality for detection	7
2.2.1 Varying the amount of noise	8
2.2.2 Varying the number of noisy images	8
2.3 Annotation quality for classification	8
2.4 Number of training images	12
2.4.1 Classification	12
2.4.2 Detection	12

3 Neural Network Architecture

15

3.1 Softmax vs Sigmoid activation for classification	15
3.2 Adding center loss	16
3.2.1 For fine-tuning	16
3.2.2 For training from scratch	18
3.3 Testing ResNet	18
3.4 Using larger input images	19
3.5 Sharing a convolutional model for multi-domain classification	20

2

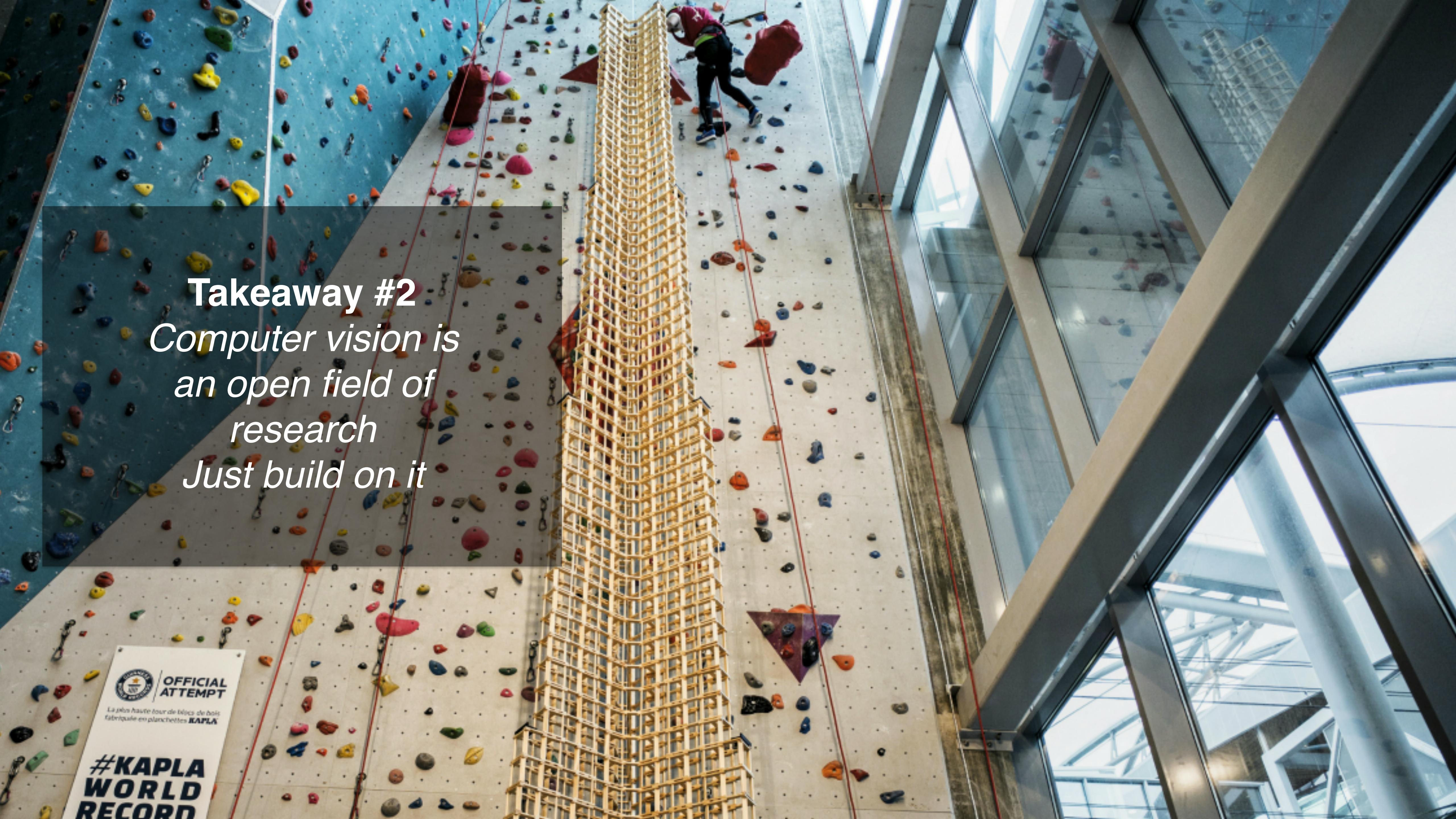
CONTENTS

3.6 New architectures for Faster-RCNN	23
3.6.1 ResNet101 and Inception-ResNetv2 for feature extraction	23
3.6.2 Using negative samples in training data	25
3.6.3 Lowering the number of sampled boxes at training time	26

4 Exploring training with deep features

28

4.1 Finding the best features	28
4.2 Fine-tuned neural network vs deep-featured SVM	29
4.3 Hyperparameter variation for Logistic Regression	29

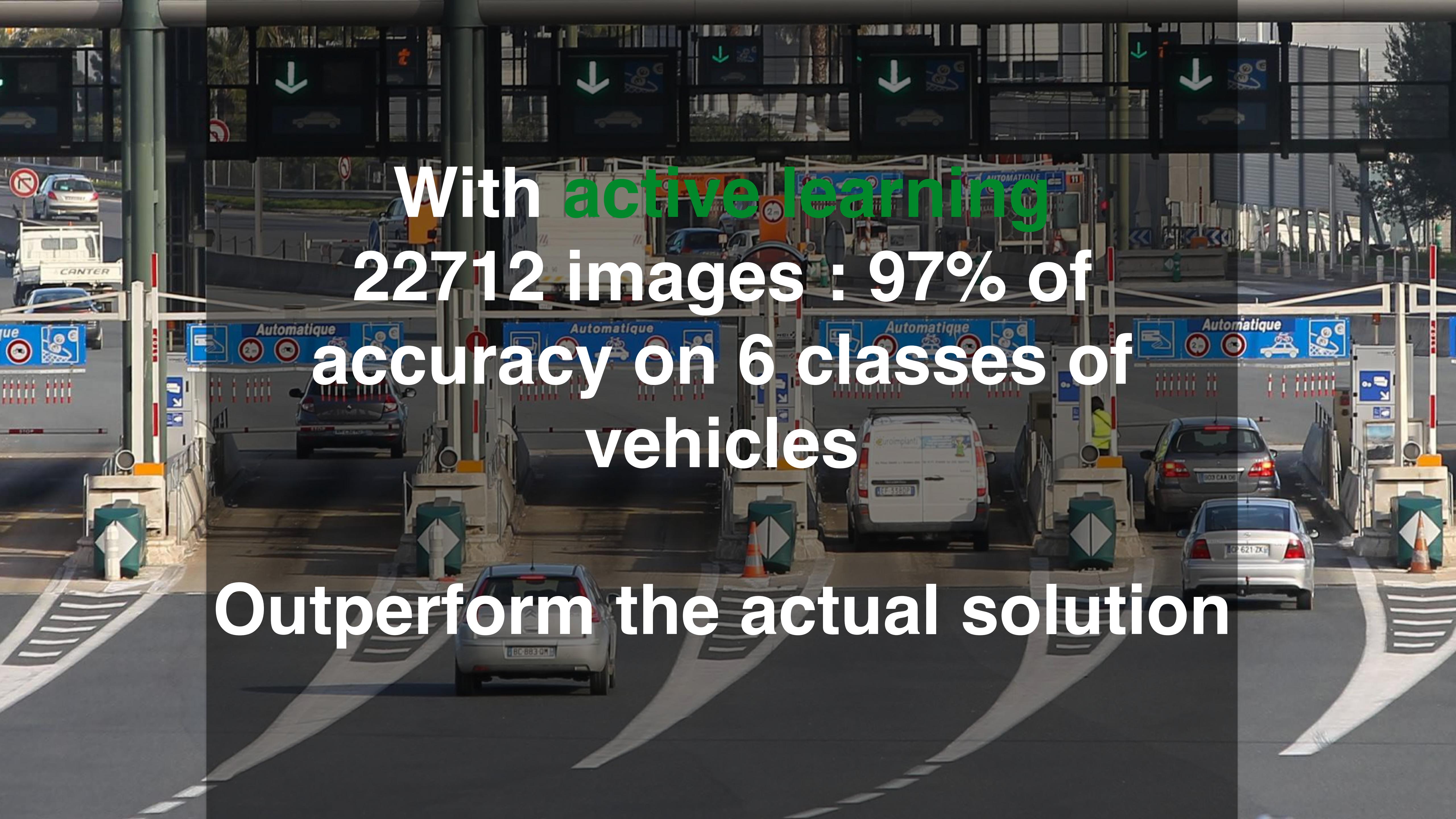


Takeaway #2
*Computer vision is
an open field of
research
Just build on it*



La plus haute tour de blocs de bois
fabriquée en planchettes KAPLA

**#KAPLA
WORLD
RECORD**

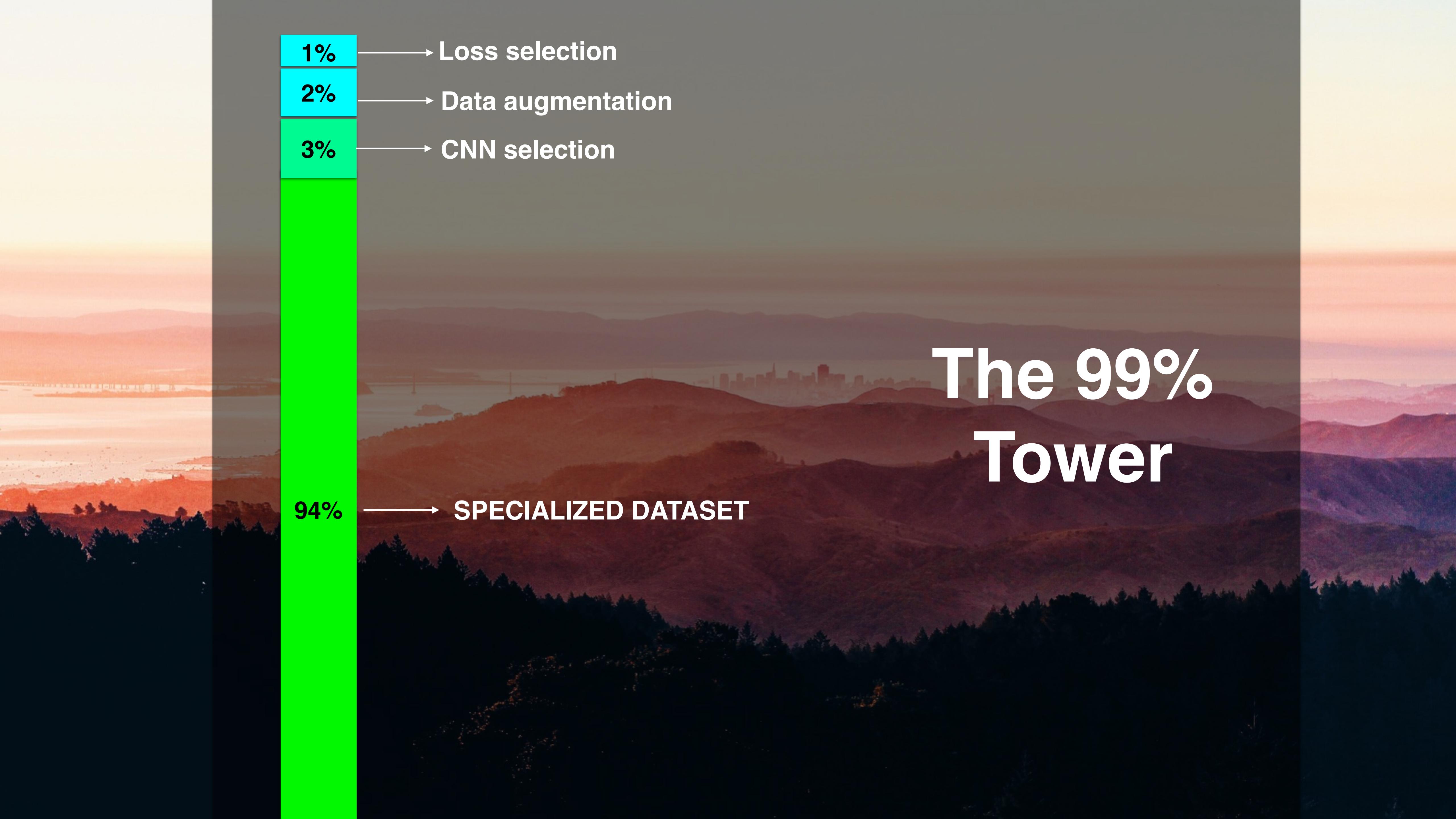


With **active learning**
22712 images: 97% of
accuracy on 6 classes of
vehicles

Outperform the actual solution

Takeaway#3

Specialized
computer vision
applications are
thrifty on data



1%

→ Loss selection
2%

3%

→ Data augmentation

3%

→ CNN selection

94%

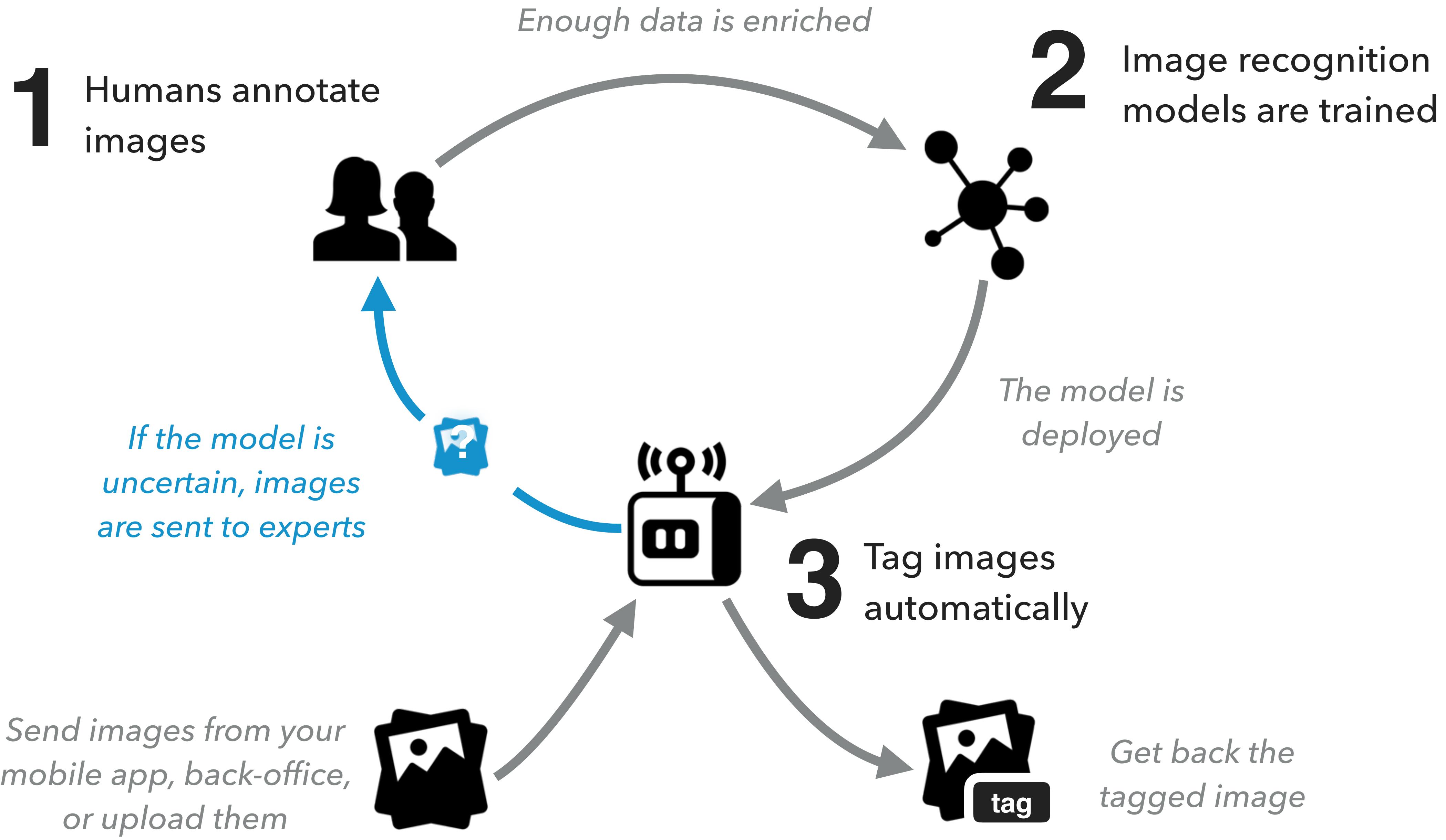
→ SPECIALIZED DATASET

The 99% Tower



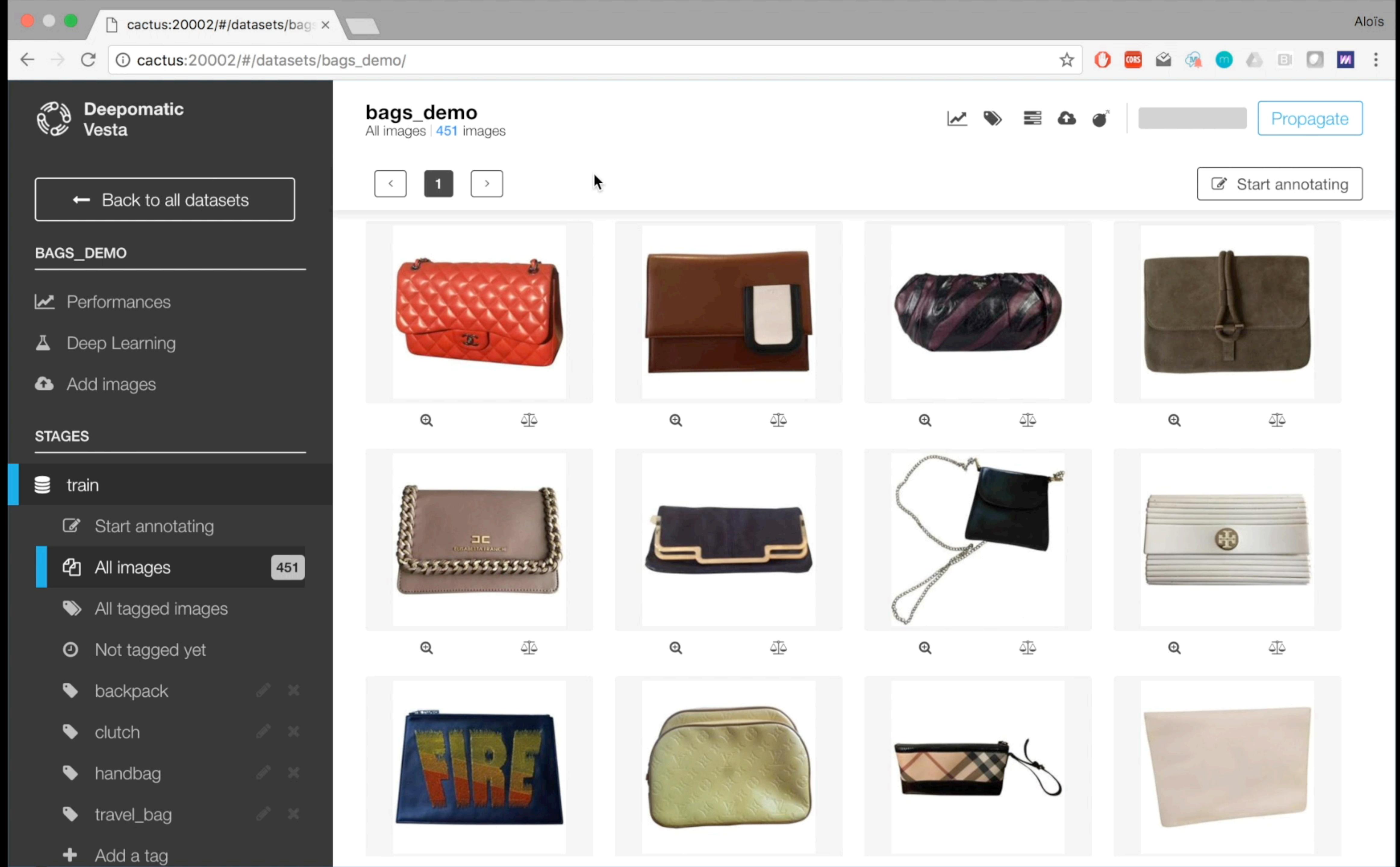
Takeaway#4

A good dataset is evolving
human-machine interactions
maintain and increase
performances



How to build a powerful dataset ?

- 1) Use your own images or videos
- 2) Have a very very strong quality review process
- 3) Active learning maximize AI performances and minimize the annotation effort
- 4) A dataset is not something frozen - humans must be in the loop to improve it over time



Thank you for your time

Augustin Marty
augustin@deepomatic.com
+33 6 51 86 87 21



1 Humans annotate images



Enough data is enriched

2

Image recognition models are trained



If the model is uncertain, images are sent to experts



Send images from your mobile app, back-office, or upload them

3 Tag images automatically



The model is deployed

Get back the tagged image

1

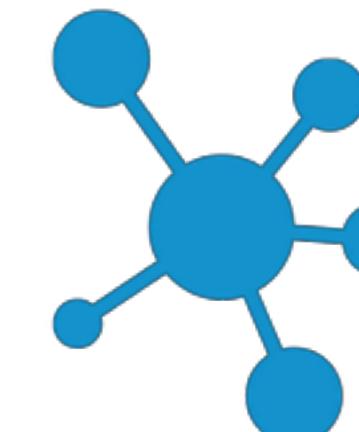
Humans annotate images



Enough data is enriched

2

Image recognition models are trained



If the model is uncertain, images are sent to experts



The model is deployed

3

Tag images automatically



Send images from your mobile app, back-office, or upload them



Get back the tagged image



1

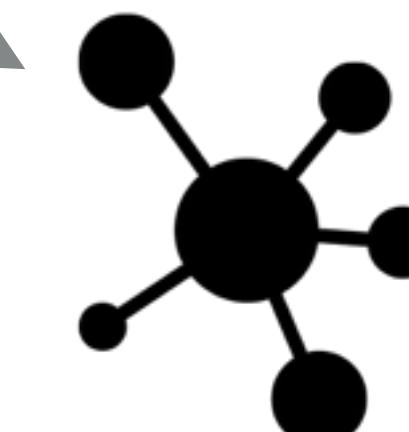
Humans annotate images



Enough data is enriched

2

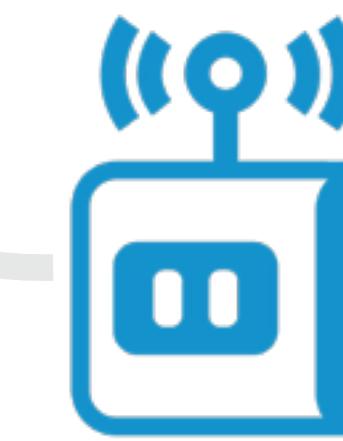
Image recognition models are trained



The model is deployed

3

Tag images automatically



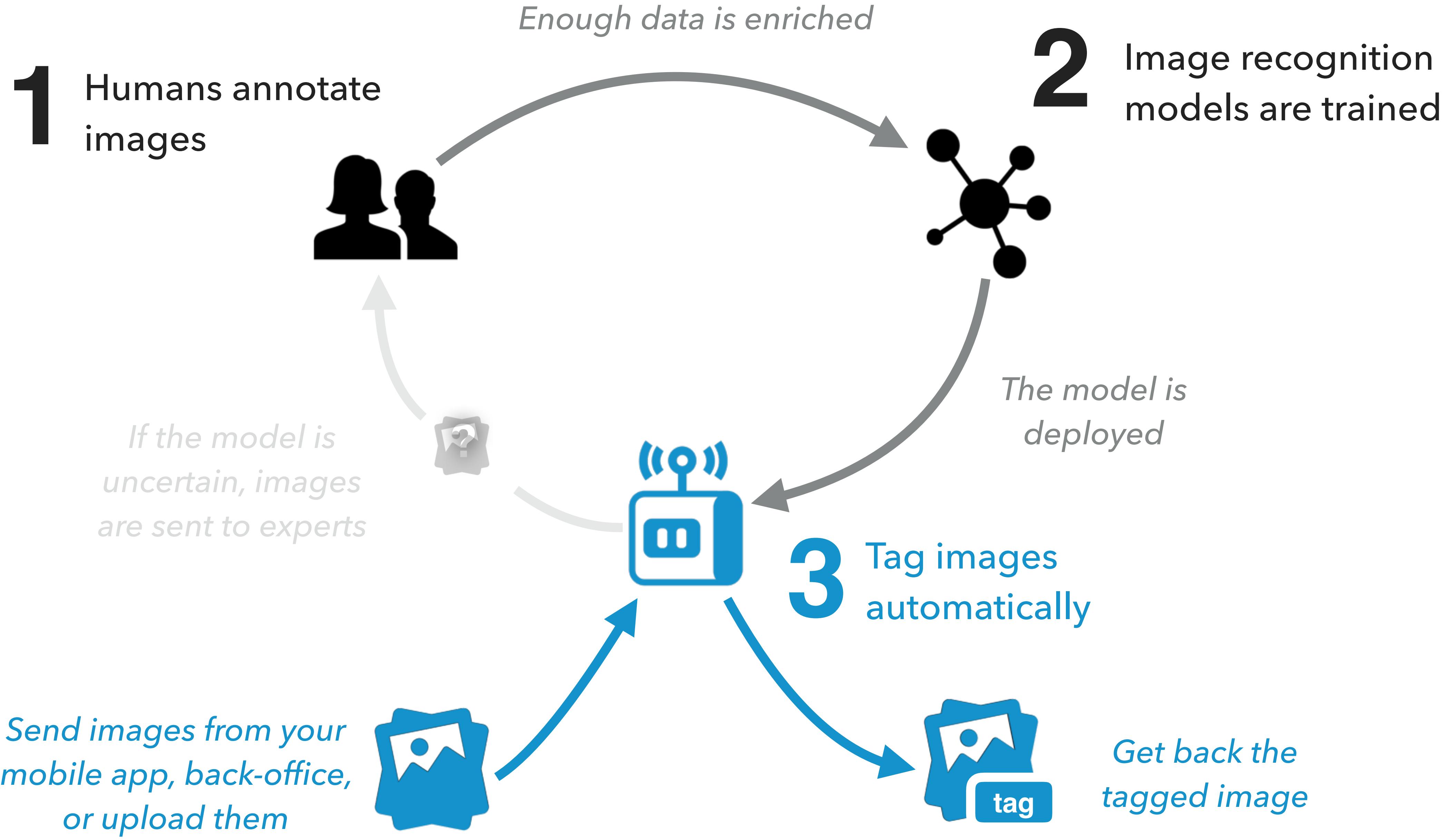
If the model is uncertain, images are sent to experts

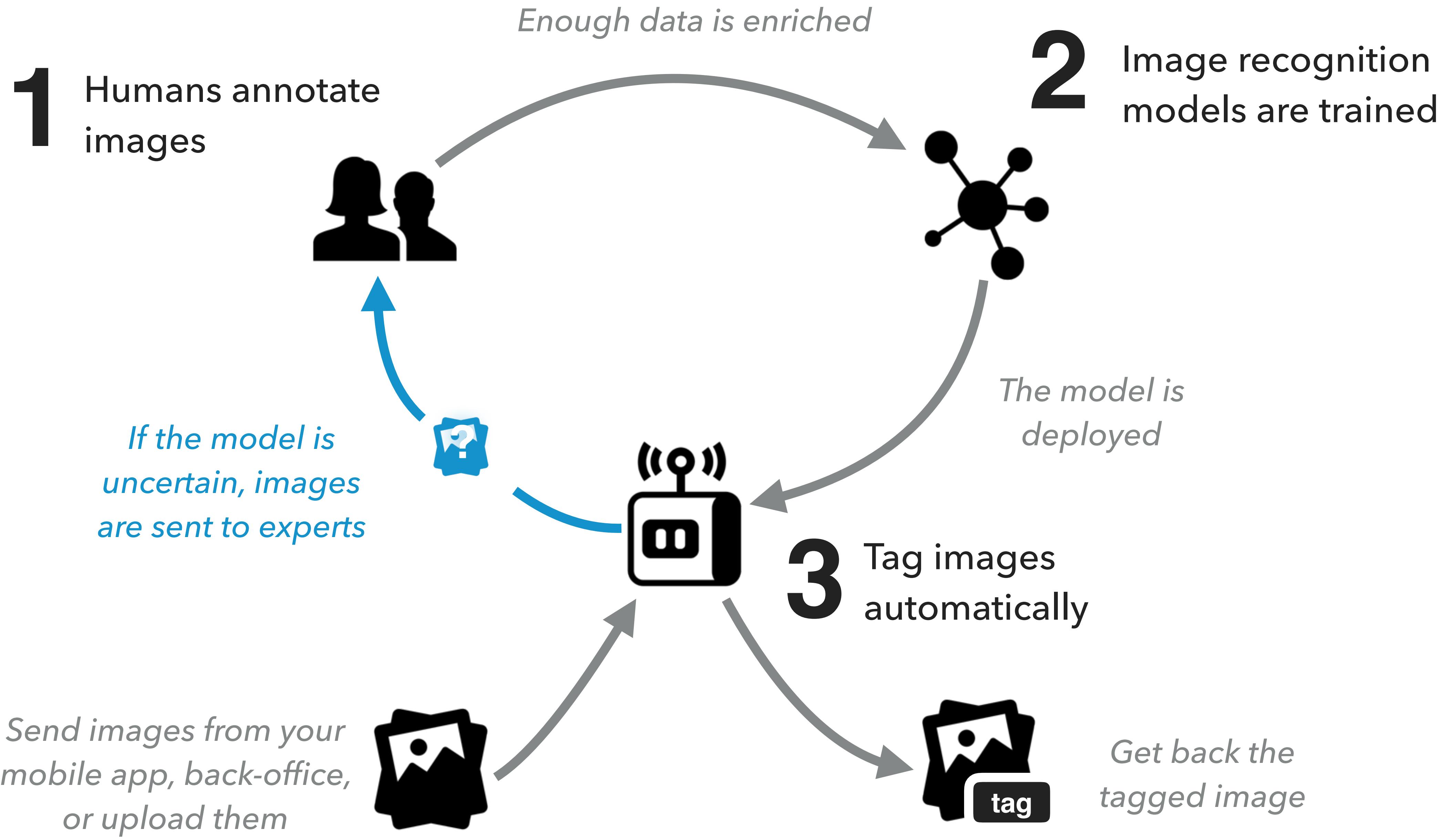


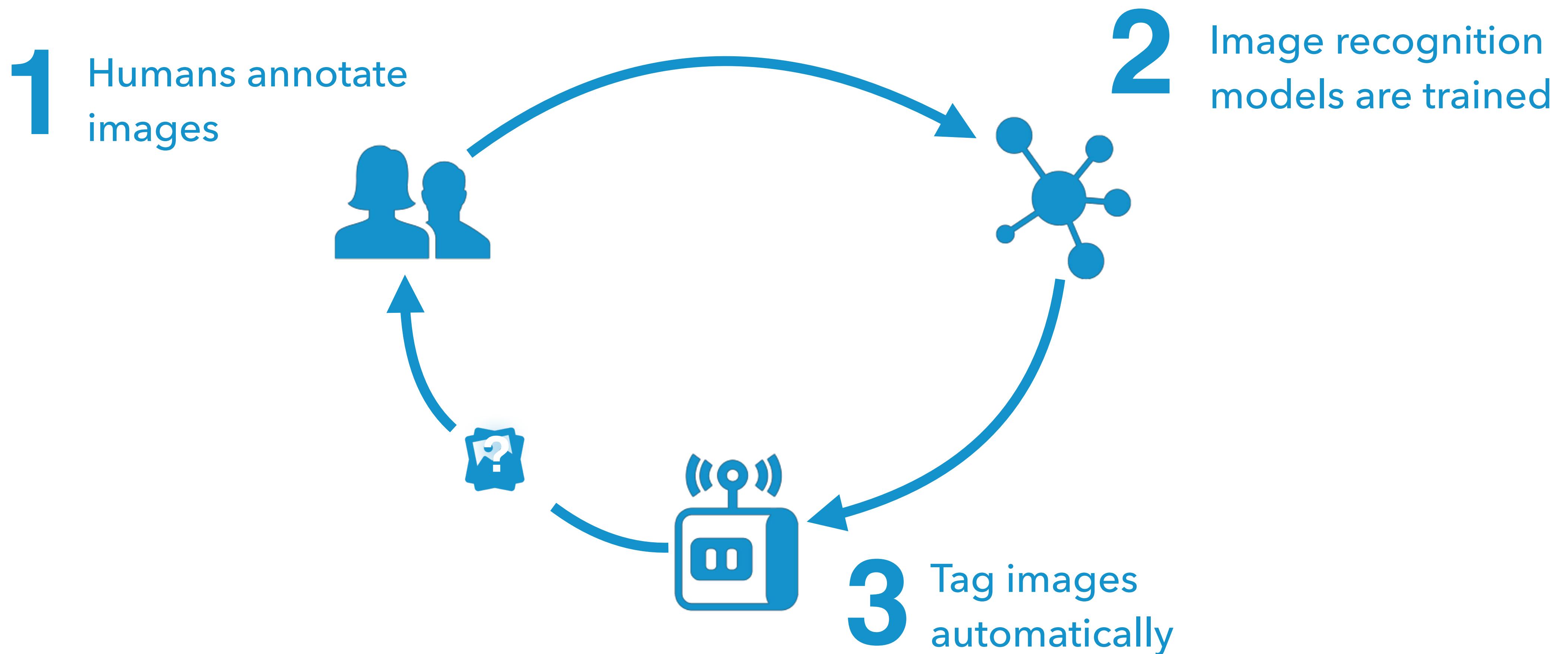
Send images from your mobile app, back-office, or upload them



Get back the tagged image







It's a *virtuous* cycle

1

Humans annotate
images



Enough data is enriched

2

Image recognition
models are trained



Models improve over time and let experts handle difficult tasks

*If the models
are uncertain, images
are sent to experts*

*The models
tag images
automatically*

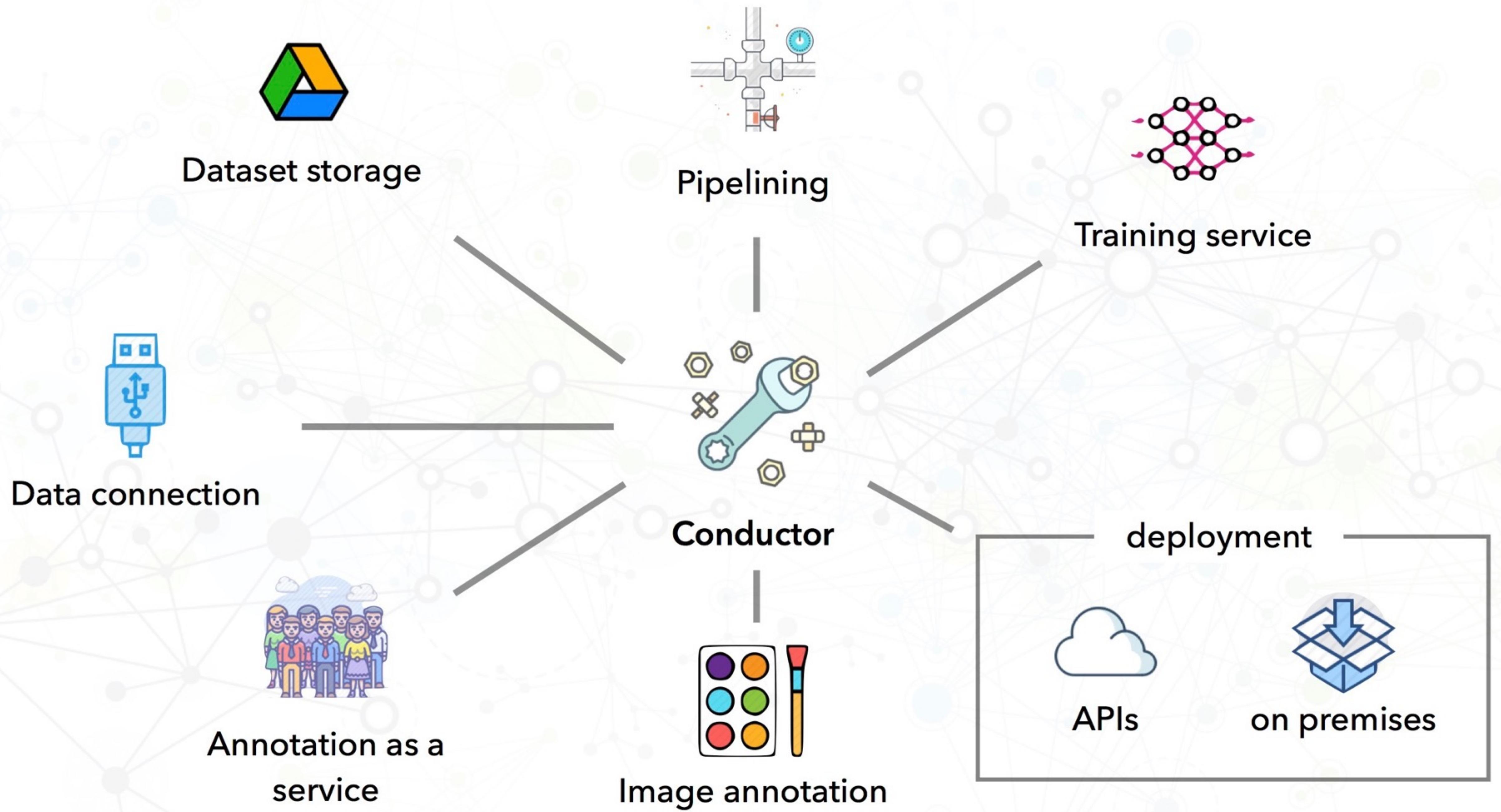
*Send images from your
mobile app, back-office,
or upload them*



*Get back the
tagged image*



VESTA





Dataset storage

- Store datasets in a structured way
- In the future: allow for mutualization of the datasets
- Import / export capabilities



Data connection



Conductor



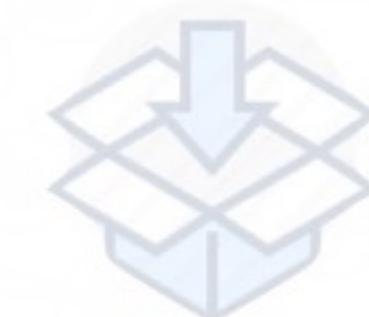
Annotation as a
service



Image annotation

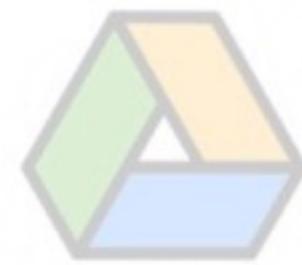


APIs



on premises

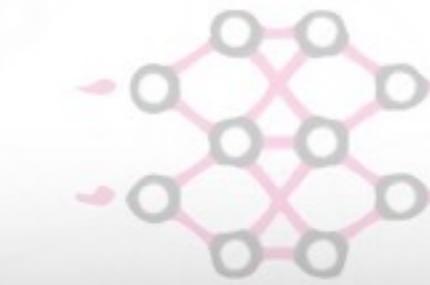
deployment



Dataset storage



Pipelining



Data connection

- Annotate images with classification / detection / segmentation information
- Annotate videos with possibly temporal annotations
- Integrate pre-annotation coming from trained models



Annotation as a
service

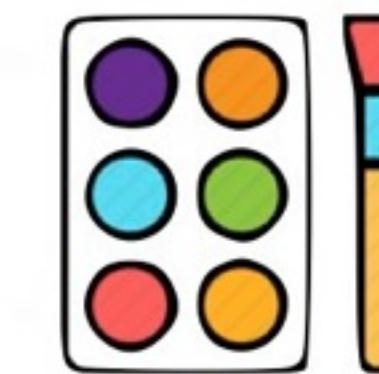
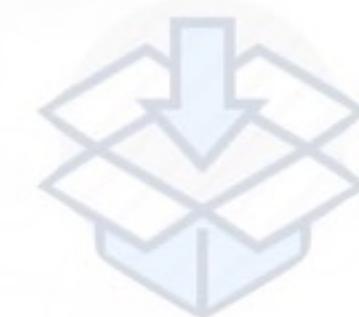


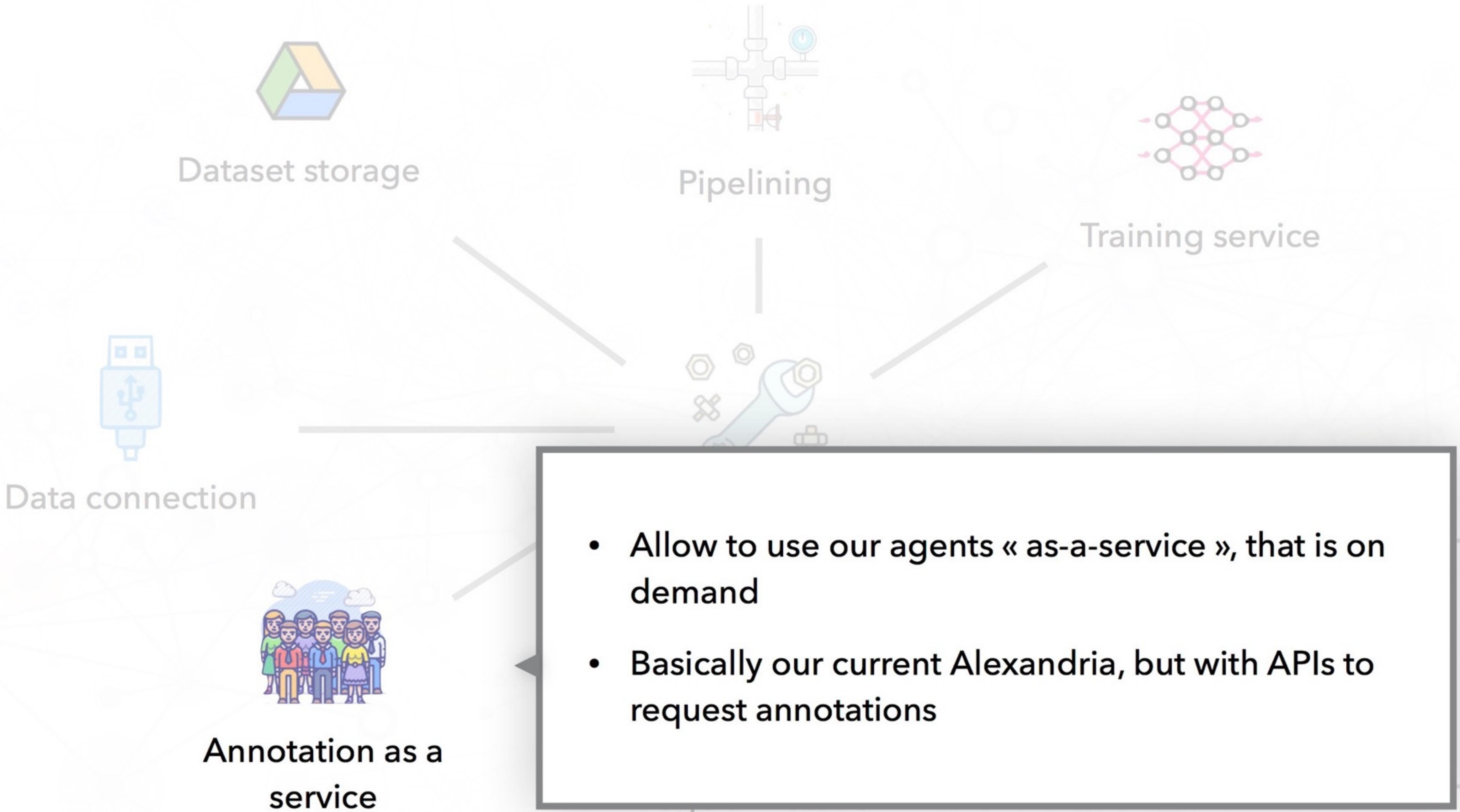
Image annotation

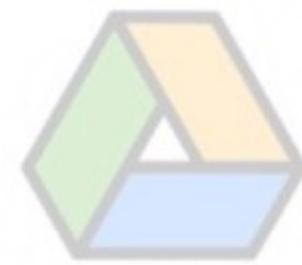


APIs



on premises

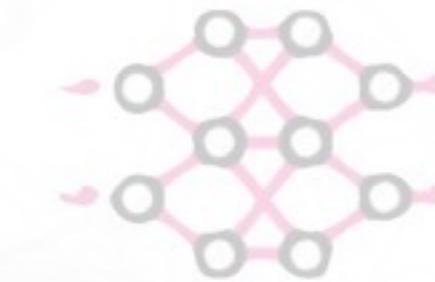




Dataset storage



Pipelining



Machine learning service



Data connection

- Connect easily with our customer data
- Crawlers (Kraken)
- Database connectors
- Interface for cameras and other sensors
- Anonymisation



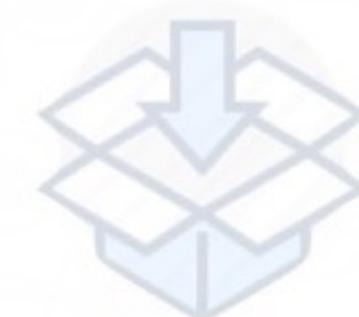
Annotation as a
service



Image annotation

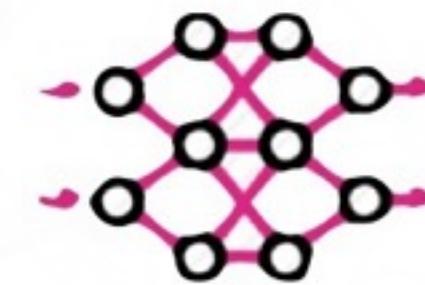


APIs



on premises

- Train models for various tasks: classification, detection, segmentation, etc.
- Use optimized strategies for training (hyperparameters tuning, data augmentation, etc.)
- Integrate our knowledge of what works, what doesn't work, etc.
- Possibly: model compression for constrained environments



Training service



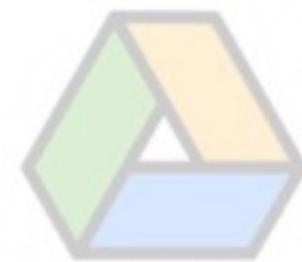
deployment



on premises

Annotation as a
service

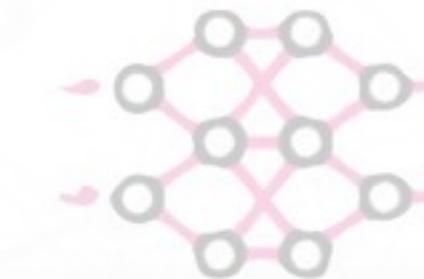
Image annotation



Dataset storage

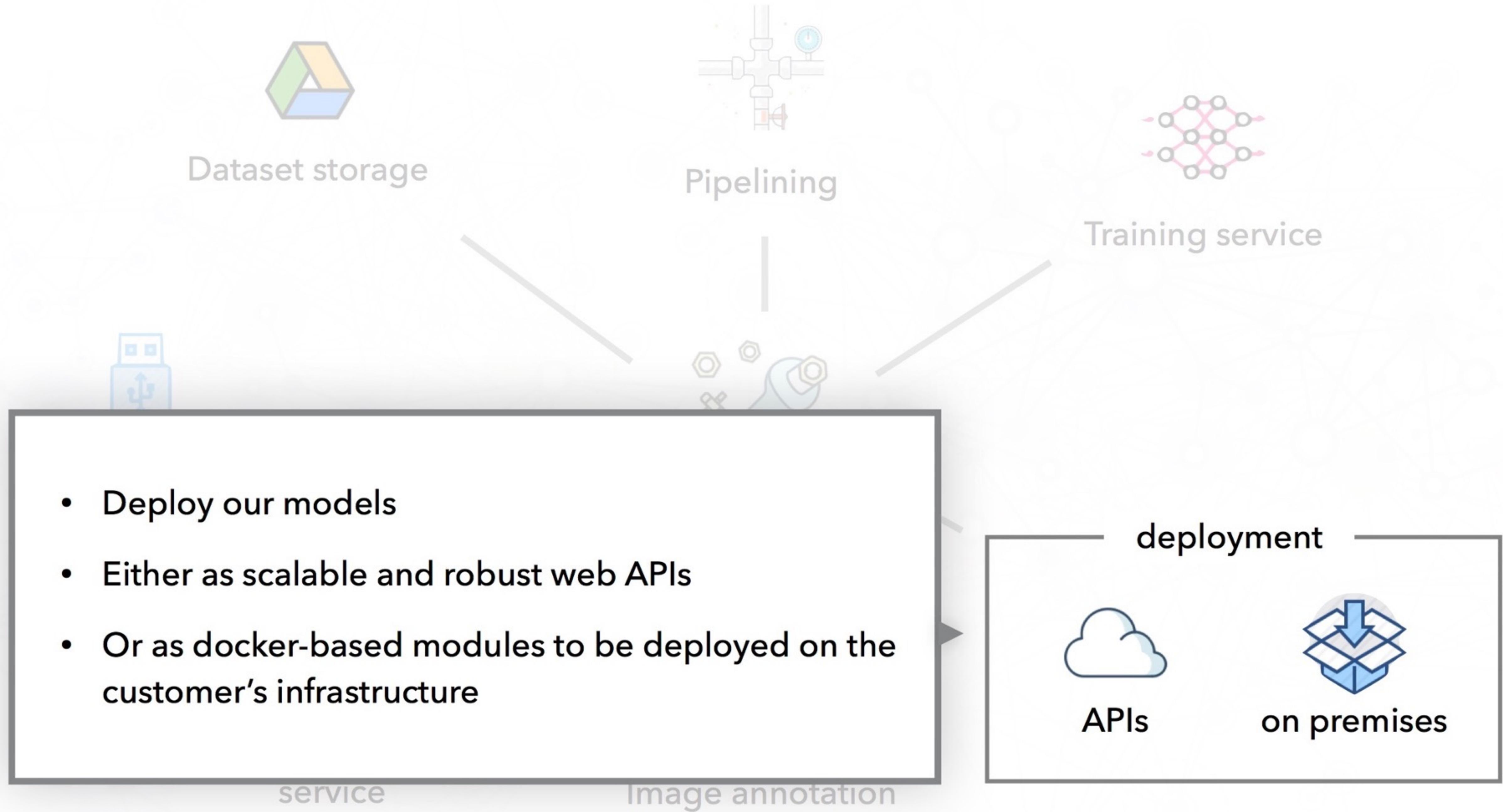


Pipelining



Dat

- Allow to build pipelines using image recognition. For instance:
 1. Take an image, detect various clothes
 2. Use specific classifiers on each of the boxes, to predict fine-grained attributes.
- Design recognition scenarios
 - ex: « a human is in front of a dinosaur » or « a red car is outside the road ».
- Use additional pre and post-treatment
 - ex: tracking, video processing, etc.





- A module to orchestrate and connect all other modules
- Basically what's currently called **vesta** today
- Used either by:
 - us to solve a customer problem
 - the customer directly

Annotation as a
service



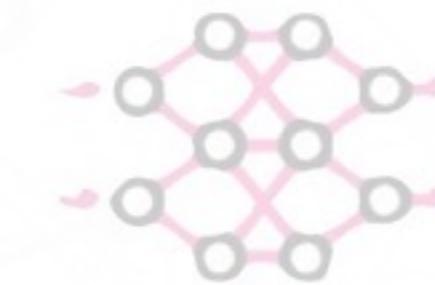
Pipelining



Conductor



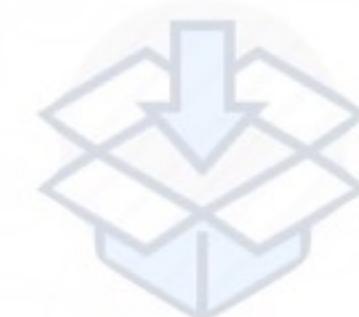
Image annotation



Training service



APIs



on premises

