Automated Neural Network Model Training: The Impact on Deploying and Scaling ML at the Edge

Tim Hartley

VP Product & Marketing

SeeChange Technologies Ltd

SeeWare Platform

Smart Streets

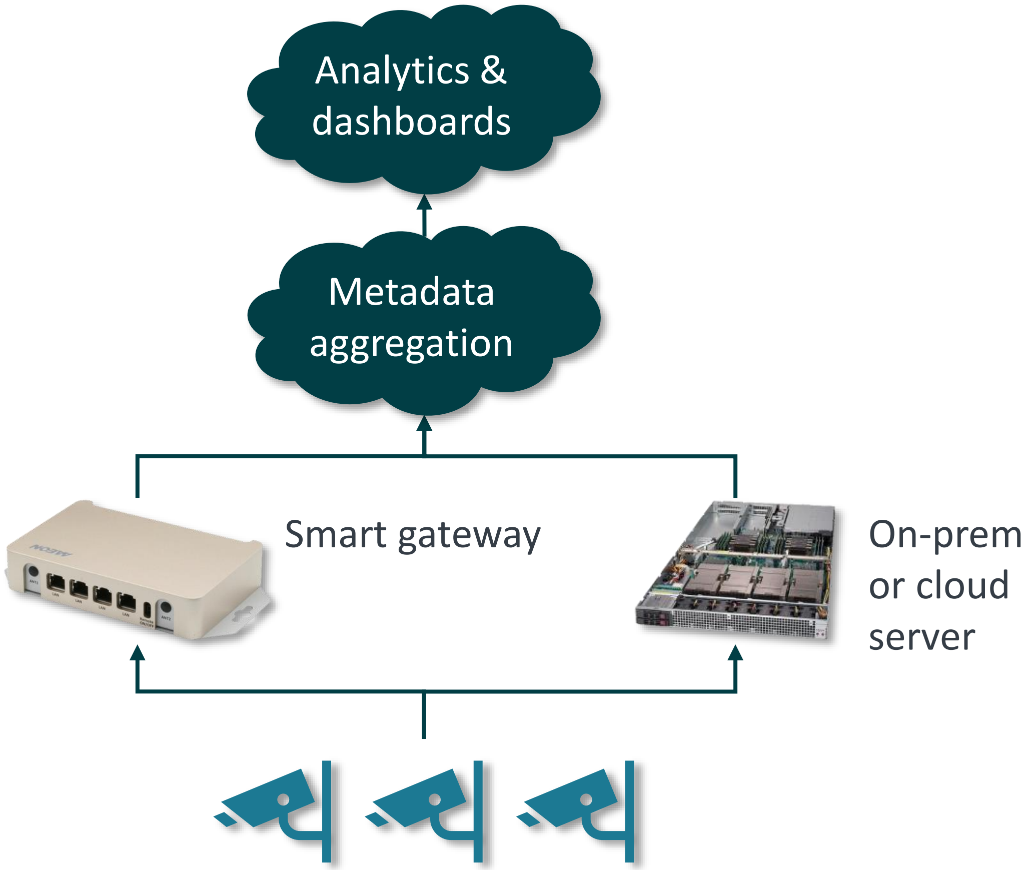
Retail

Smart Buildings

Healthcare

Traffic Management

Edge to Cloud: The Challenge now is Deployment



Edge-to-cloud model for CV applications

- Multiple sensors: typically cameras
- Local smart gateways / on-prem servers
- Running detection models & streaming insight metadata into the cloud
- Cloud aggregation with analytics & dashboards

The challenge now is how you scale

- Many real-world use cases need models to be built or tuned in real time

We're going to look at a two examples of how we can start to **automate** model creation & tuning

Federated Learning?

The typical understanding of federated learning:

- Spreading a model training pipeline across multiple edge devices

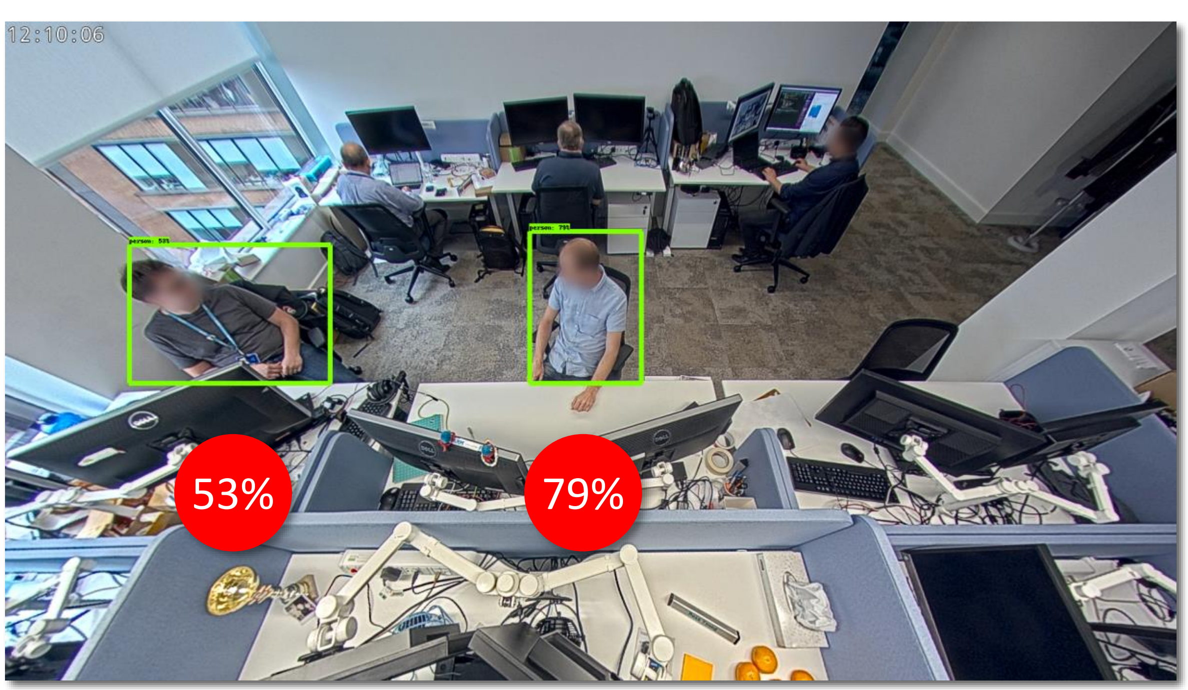
This talk is more accurately looking at learning from federated training data

- Cloud aggregating training data gathered from multiple edge devices to train / tune models that can be redeployed back to the edge via over-the-air updates

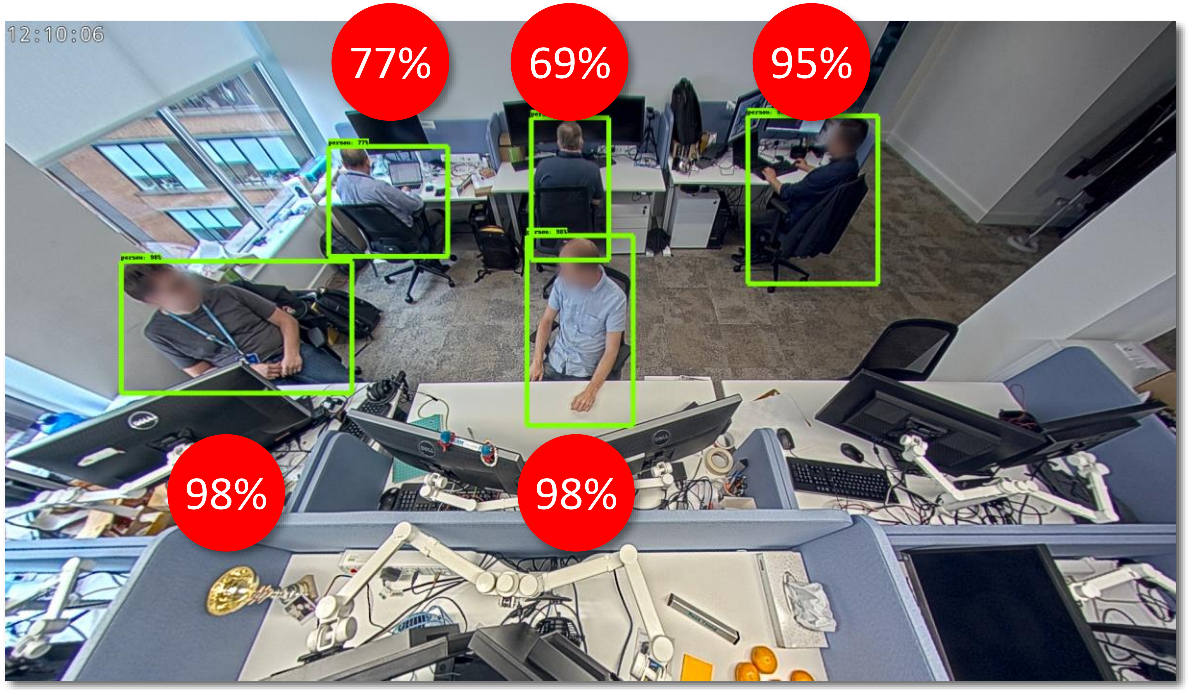
Example 1: Auto Model Tuning

Comparison of Detection Performance

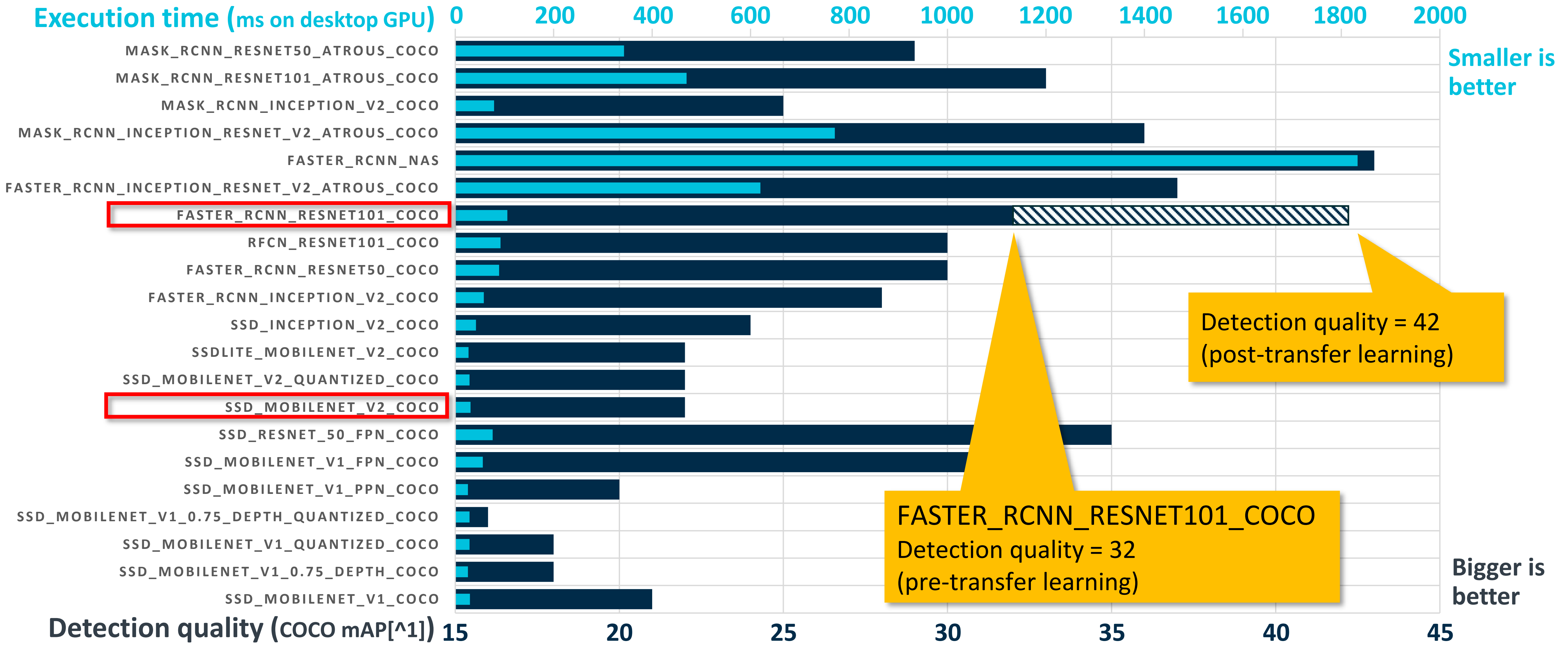
SSD_MOBILENET_V2_COCO



FASTER_RCNN_RESNET101_COCO

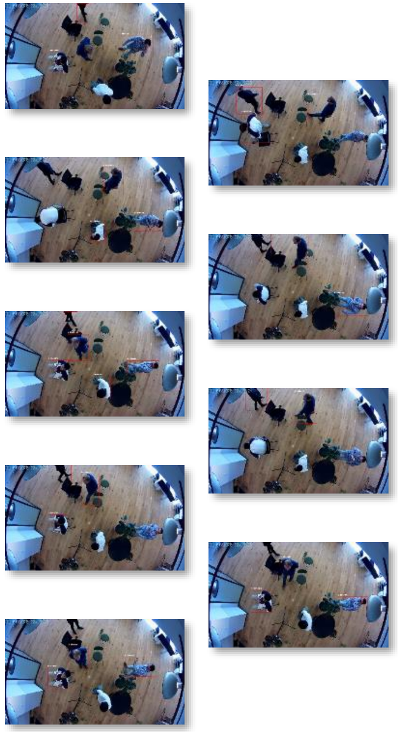


Impact of Transfer Learning



SeeChange an arm company

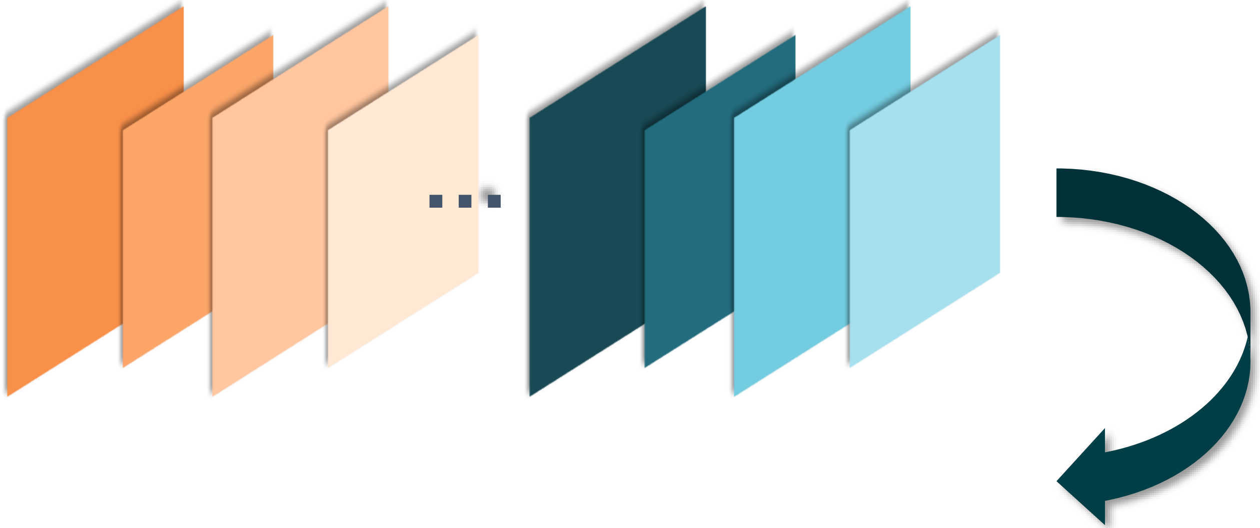
Tuning the Model Using Transfer Learning



Gather and label new set of sample images (~250)

Crop & create rotated, brightness, quality variations

Pre-trained model based on FasterRCNN & ResNet101



Feed through training pipeline using existing network

Repeat for each image

Transfer Learning for Model Tuning. Does it Scale?

In practice, what does model tuning involve?

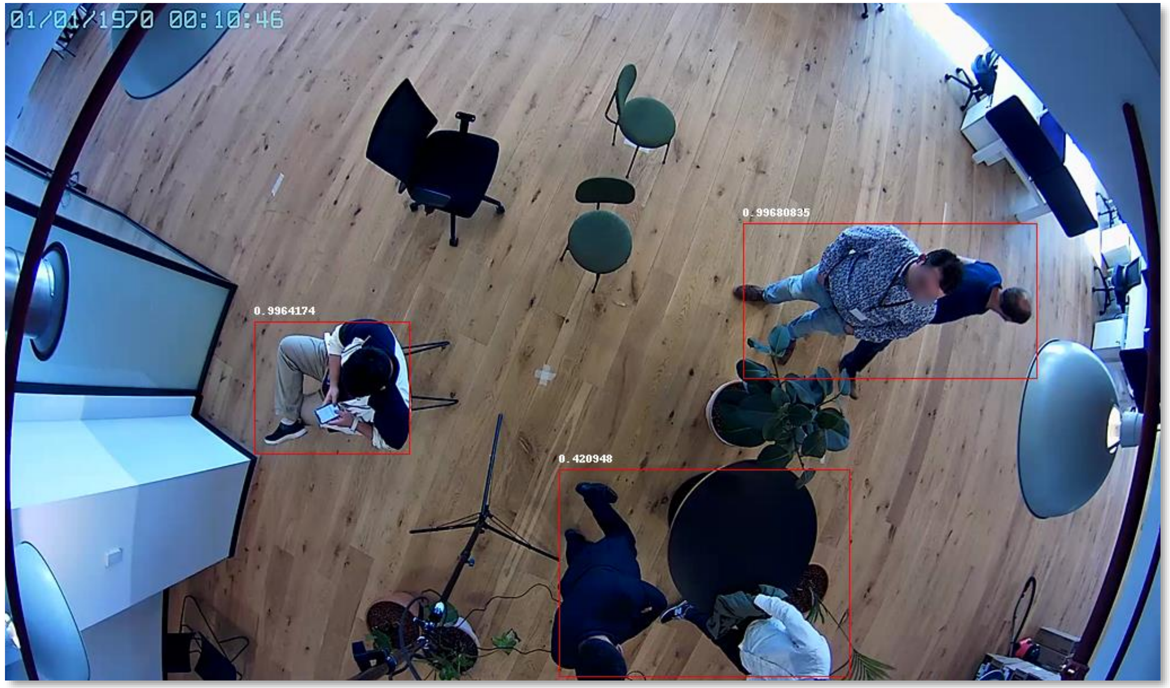
- Gathering training data (100s to many 1000s of images)
- Image labelling
- Running the training process

Bottom line: model training is a time-consuming, laborious process

- Difficult to scale to multiple locations
- Particularly if the tuning is a requirement to take in conditions specific to each location

For many use cases this just won't be viable

- Unless you can automate the process, ROI is going to be hard



Multiple Thresholds

Set detection threshold T1

For detections >= T1, assume **true positive**
For detections < T1, assume **true negative**

e.g. T1 = 75%...

T1

Discard | Insight

Set additional threshold T2, where T2 < T1

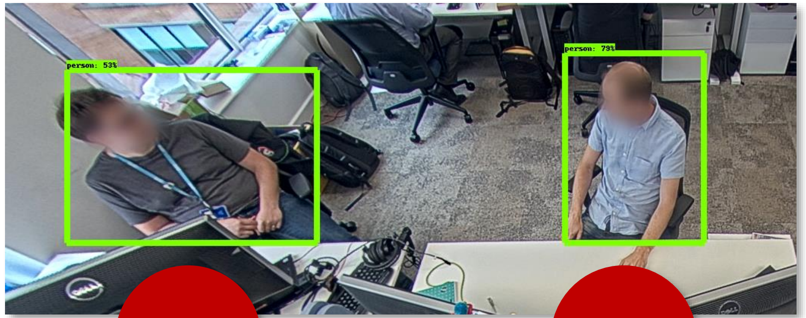
For detections < T2, assume **true negative**
For detections between T1 & T2, assume **item of interest**

e.g. T1 = 75%, T2 = 50%...

T2 T1

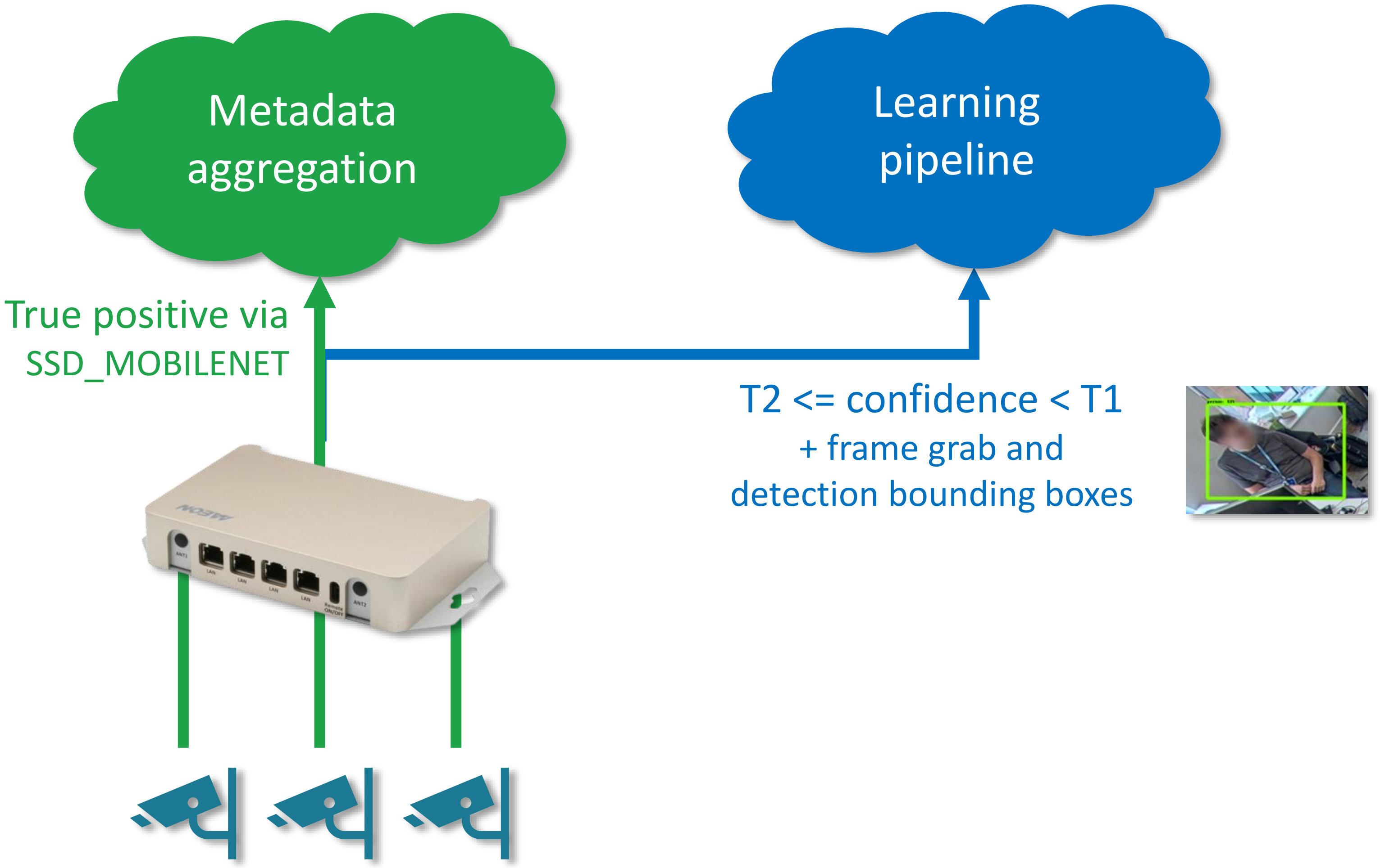
Discard | ?? | Insight

SSD_MOBILENET_V2



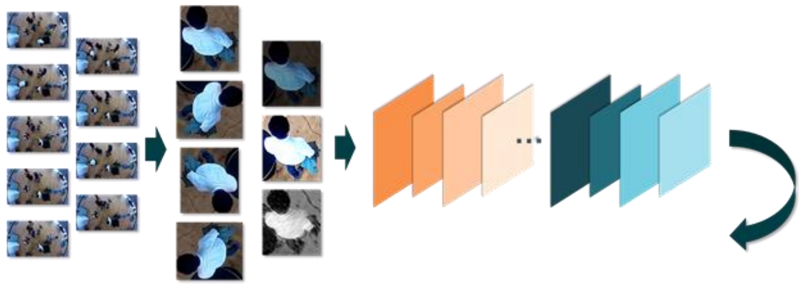
53% 79%

Tuning / Improving the Edge Model



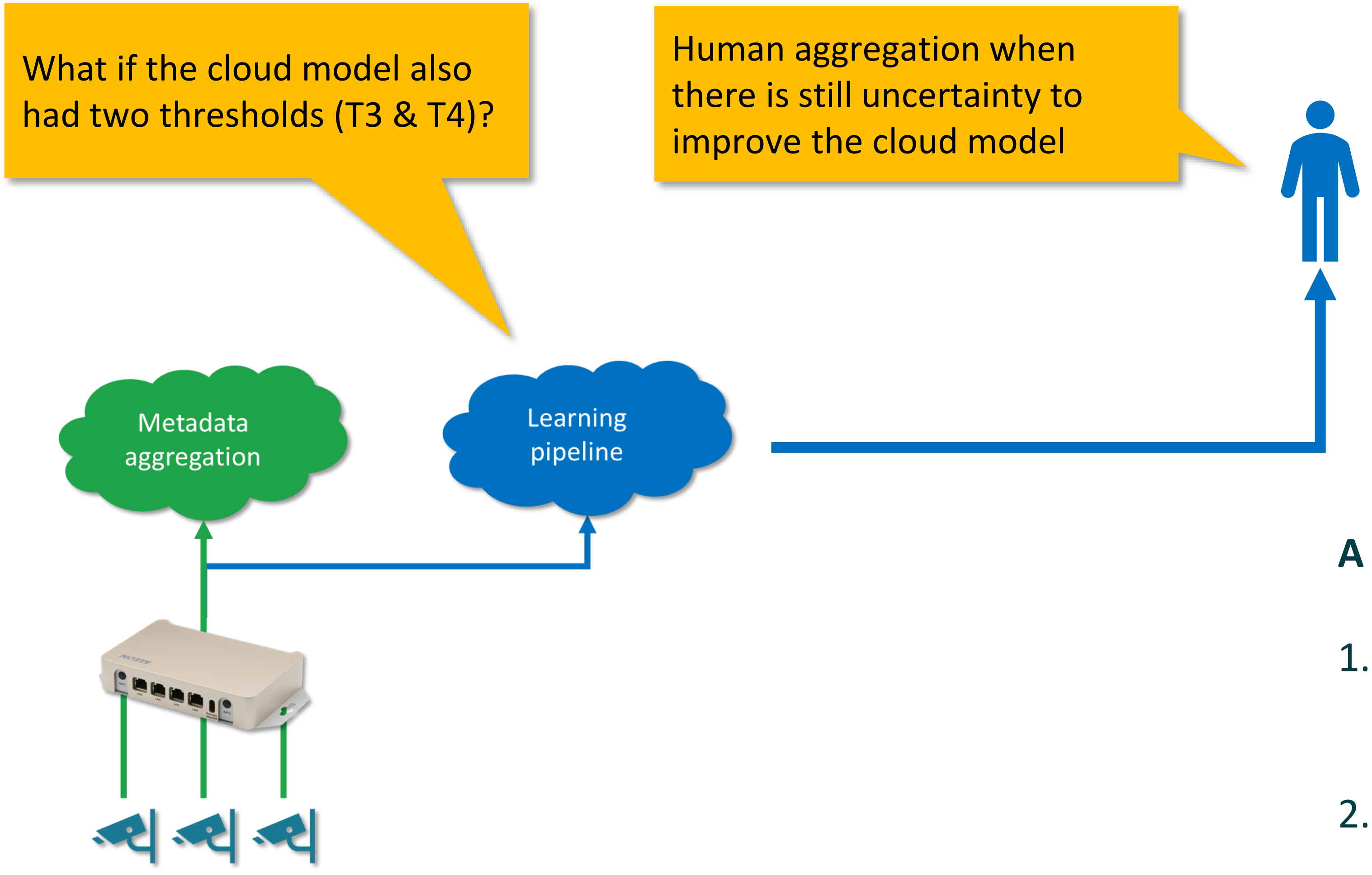
Learning Pipeline

1. Put images received through cloud model (e.g. **FASTER_RCNN_NAS**)
2. If detections above set threshold T3, then add image to true +ve list
3. Periodically, use transfer learning pipeline to tune the edge model with the collected true +ve images



4. OTA update model back to edge gateways

Could We Go One Step Further?



T4 <= confidence < T3
+ frame grab and
detection bounding boxes

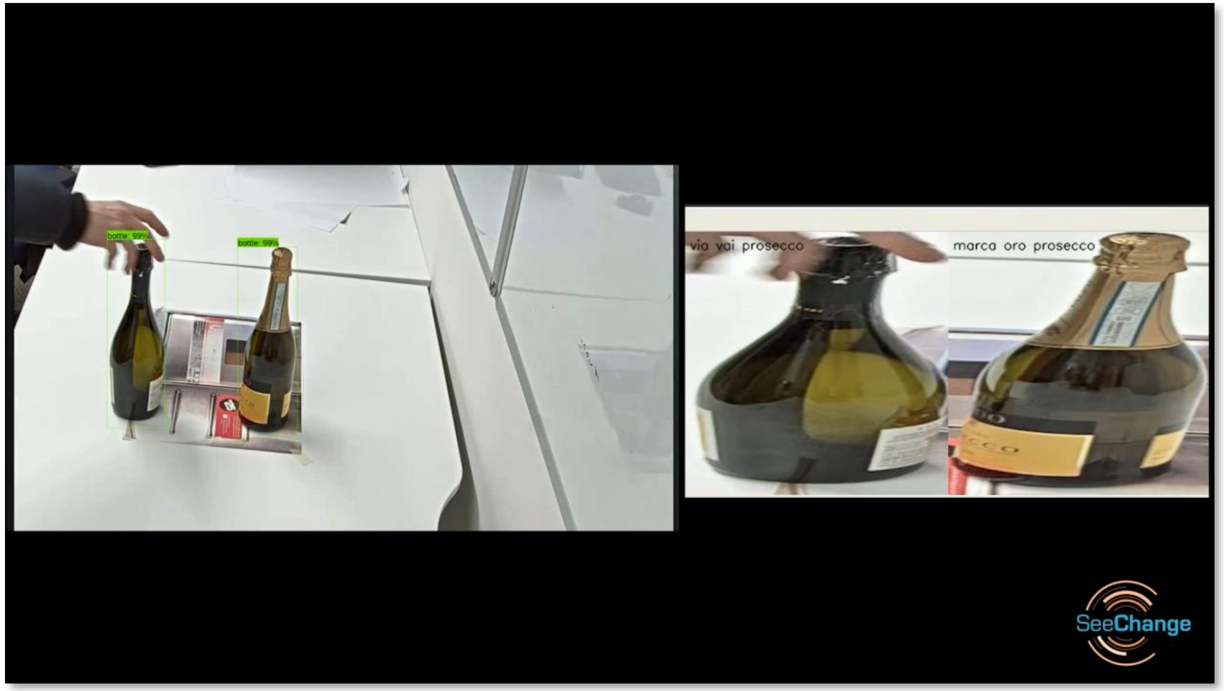
A cascade of re-training

1. Human aggregation on a relatively small number of images helps to tune the cloud model
2. The cloud model helps to tune the edge model

SeeChange an arm company

Example 2: Zero Touch Model Learning & Tuning

Real Time Product Recognition at Retail Store Check Out



https://youtu.be/S6tAprp-bUU

Helping reduce retail product shrink

Reconcile two lists

- What is seen vs what is scanned
- And alert if there is a discrepancy

Uses a standard object recognition model

- Product recognition works well

But how does this scale?

- Who trains the model?
- What happens when product packaging changes?
- Serious risk in reduced ROI if this process cannot be automated

Training for New Products: The Manual Approach

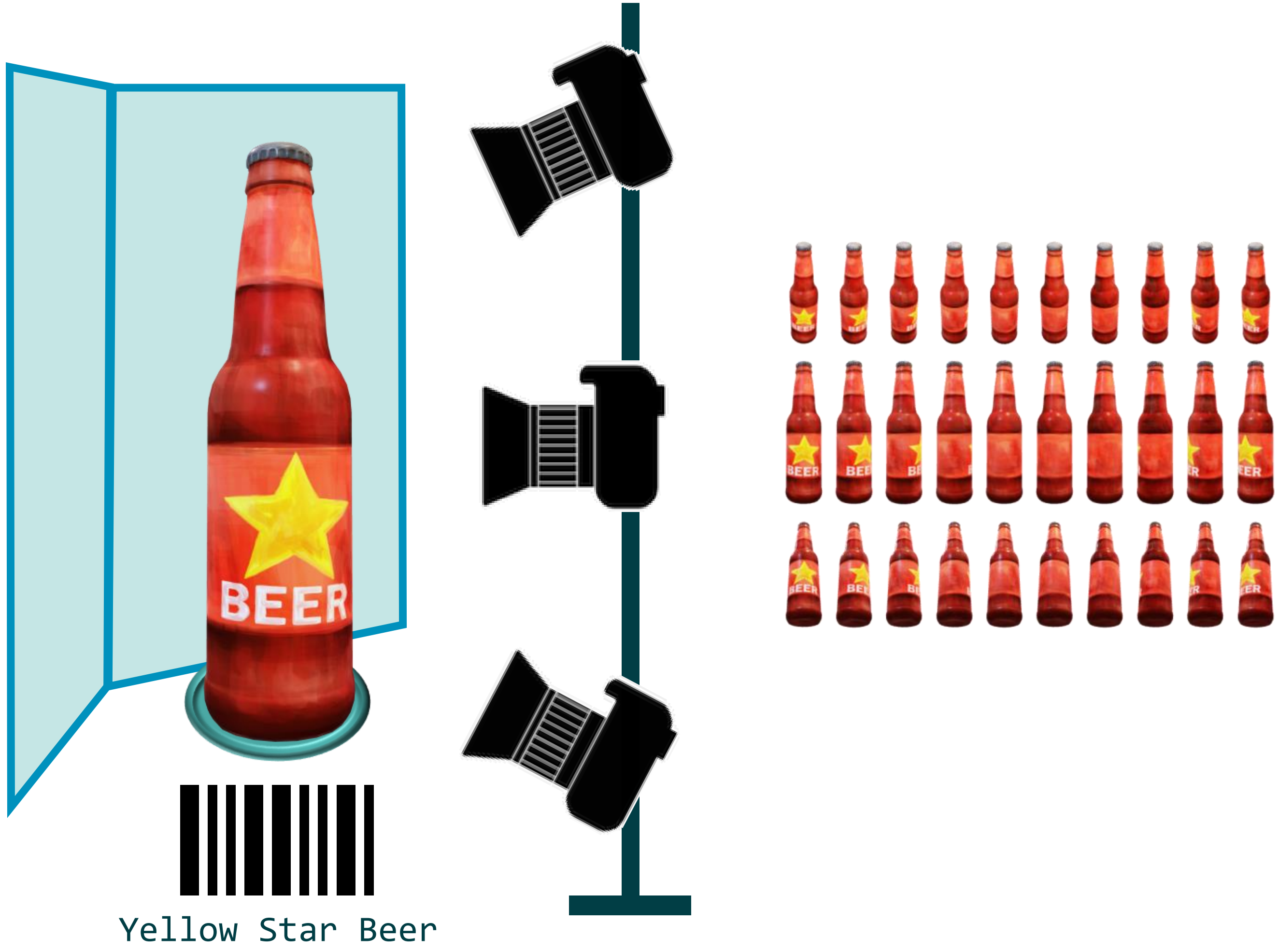
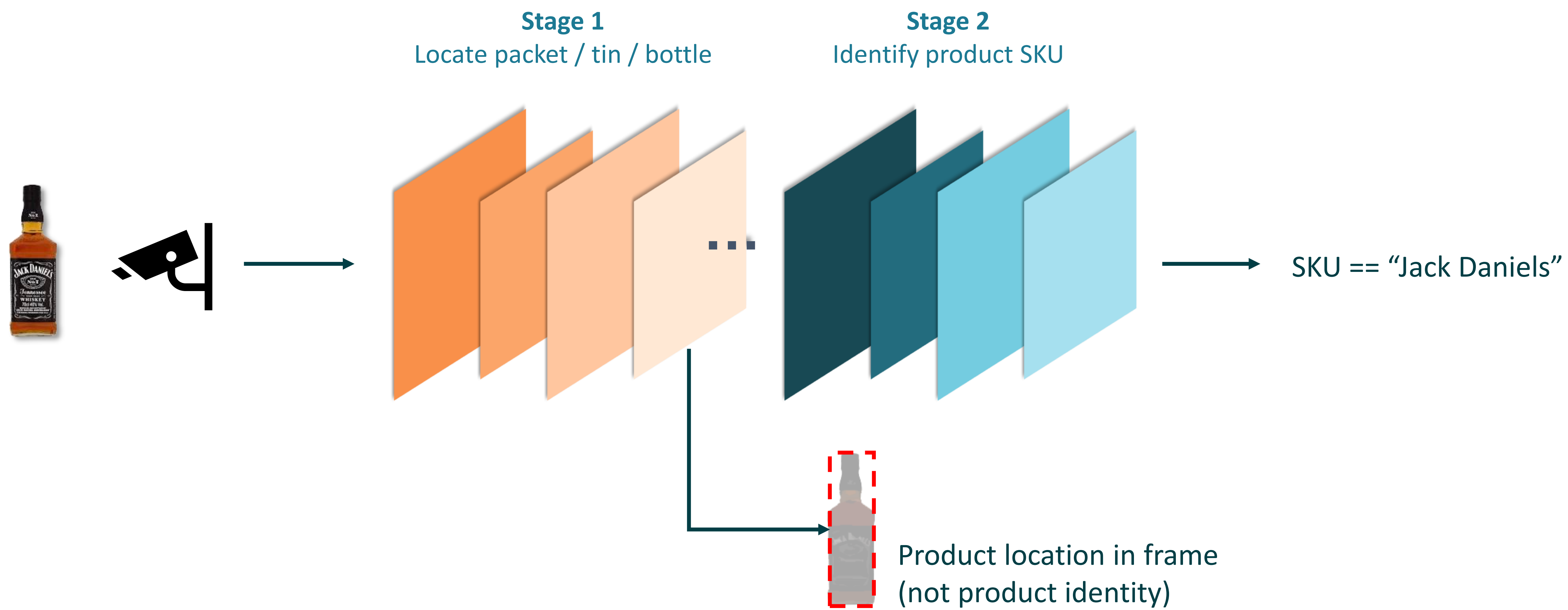


Image capture pipeline:

- Scan the product's barcode to register it with the existing stock database
- Put product into capture area
- Rotate the product whilst taking pictures
- Using transfer learning, the images retrain the machine learning model so the new product can be recognised

Automation: Getting More from your Model Pipeline

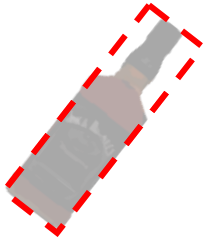


Using Honest Transactions to Train our Model

Honest transactions create ground truth we can use for product training

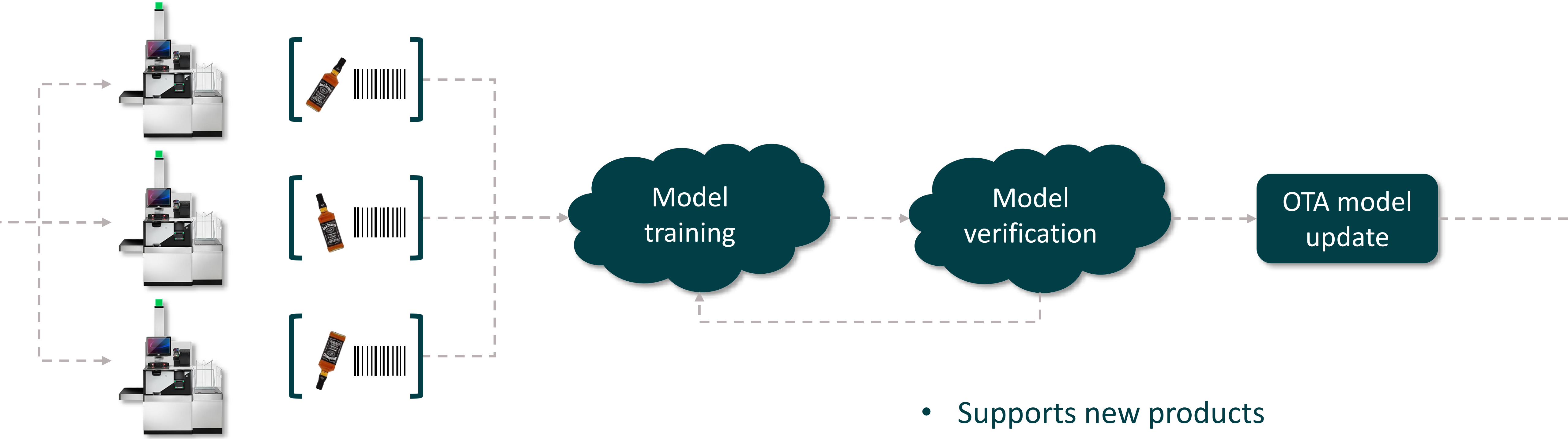
+ timestamp

With the timestamp from when something we didn't recognise was scanned we can find the associated product bounding box



Training data

Bringing Everything Together in the Cloud

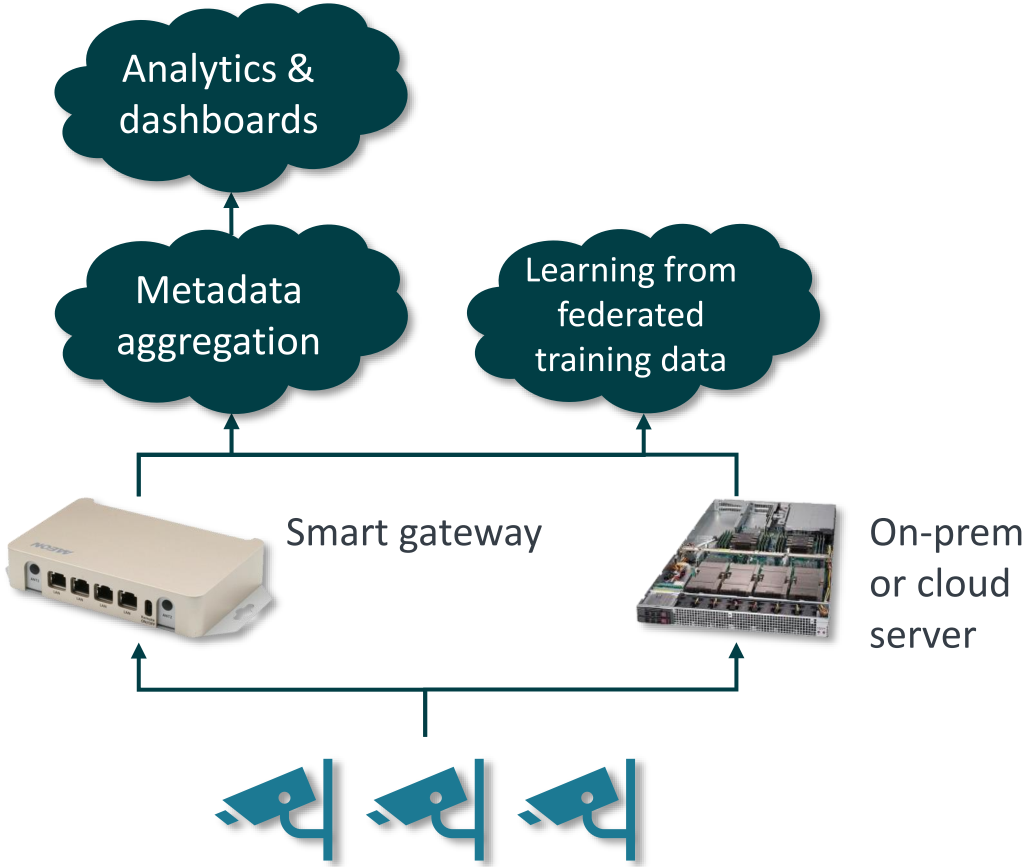


- Supports new products
- And products with updated packaging
- **<u>Zero</u>** touch for the supermarket

SeeChange
an arm company

Enabling Auto Learning: Making it Easy

Commoditizing Automated Learning



The edge-to-cloud architecture to support auto learning can be complex

- Potential additional complexity at the edge
- Cloud infrastructure to handle incoming training data
- In-cloud retrain & testing pipeline
- OTA deployment back to edge devices

Commoditizing these abilities is essential

- Reducing the friction for their use
- Allowing applications to leverage the significant benefits of auto-learning

Summary: Look for More at the Edge

Scaling & deployability are the new challenges

- ML and CV are becoming commoditized
- Now we need to do the same for auto-learning

Get more value from your edge cameras

- The more ground truth you can gather, the more your applications can self-learn
- This potentially challenges the design of the models we run at the edge: but the ROI payback is significant

Keep an eye on privacy

- Sending imagery into the cloud for training may effect your Data Privacy Impact Assessment (DPIA)

Example of Resource Slide

Resources

Tackling Product Recognition at Checkouts Using Neural Networks

Fanioudakis, Patel

https://seechange.ai/product-recognition-part1/

How AI Can Take The Drudgery Out Of Tuning Machine-Learning Models

(Forbes) Zeichick

https://bit.ly/2Q5Uksw

7 Jobs Humans Can Do Better Than Robots And AI

(AI won't replace soft skill jobs… do you agree?)

(SmartDataCollective) Mallon

https://bit.ly/2REWbF6

2021 Embedded Vision Summit

"IoT and Vision: Why It's a Security Minefield and How to Navigate It"

Lyndon Fawcett

SeeChange Security Architect
Wednesday 26 May, 10:30am