inovex

Let's talk about Artificial Intelligence and Robotics!

Silvia Santano and Pepper

München, 11.04.2018

Silvia Santano

- Robotic Applications developer
- Programming robots since I was 12
- At inovex since June 2016

Agenda

- Computer Vision
- Image Recognition
- Pepper
  - Softbank Robotics
  - Characteristics
  - Usecases
  - Computer Vision
  - External services
  - On-device CV with CNNs

Agenda

- Computer Vision
- Image Recognition
- Pepper
  - Softbank Robotics
  - Characteristics
  - Usecases
  - Computer Vision
  - External services
  - On-device CV with CNNs

Automatic extraction, analysis and understanding of information from images

Image: http://www.puretechsystems.com/images/ComputerVision3.jpg

Image: https://fineartamerica.com/featured/finlayson-old-house-susan-crossman-buscho.html

Image: http://www.cavareno.com/how-to-be-green-at-home/best-green-home-design-138869-2/.jpg

Image: http://wallpaperose.com/house-in-the-forest-painting.html

Image: http://www.dezineguide.com/inspiration/15-creative-exterior-houses-designs-examples/.jpg

Humans can recognize objects in images with little effort despite of huge variations

For computers, this is still a challenge…

Agenda

- Computer Vision
- Image Recognition
- Pepper
  - Softbank Robotics
  - Characteristics
  - Usecases
  - Computer Vision
  - External services
  - On-device CV with CNNs

Image Recognition

Main subfields:

› **Classification**

›› Object detection

››› Semantic segmentation

Image: http://www.coldvision.io/wp-content/uploads/2016/07/dnn_classification_vs_detection.png

Image Recognition

Main subfields:

› Classification

›› **Object detection**

››› Semantic segmentation

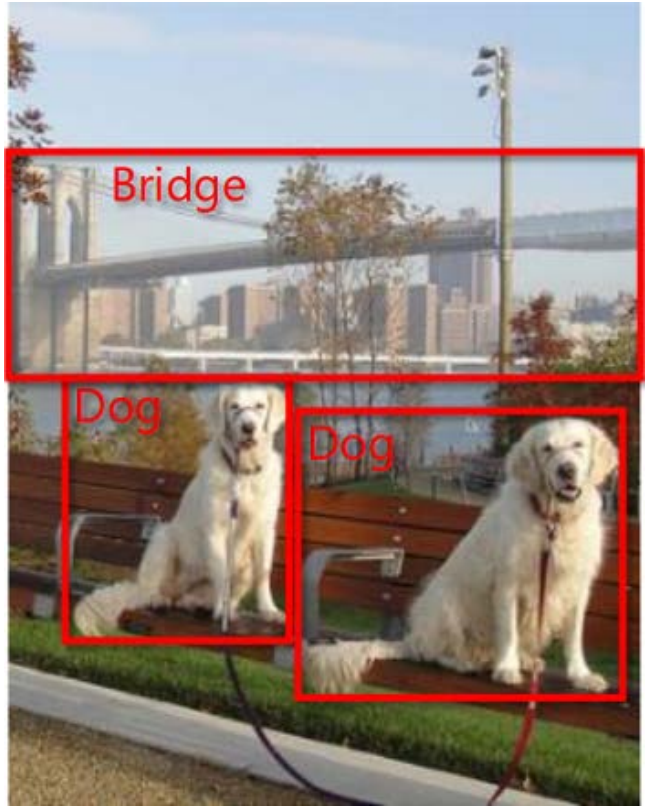


Image: http://www.coldvision.io/wp-content/uploads/2016/07/dnn_classification_vs_detection.png

Image Recognition

Main subfields:

› Classification

›› Object detection

››› **Semantic segmentation**

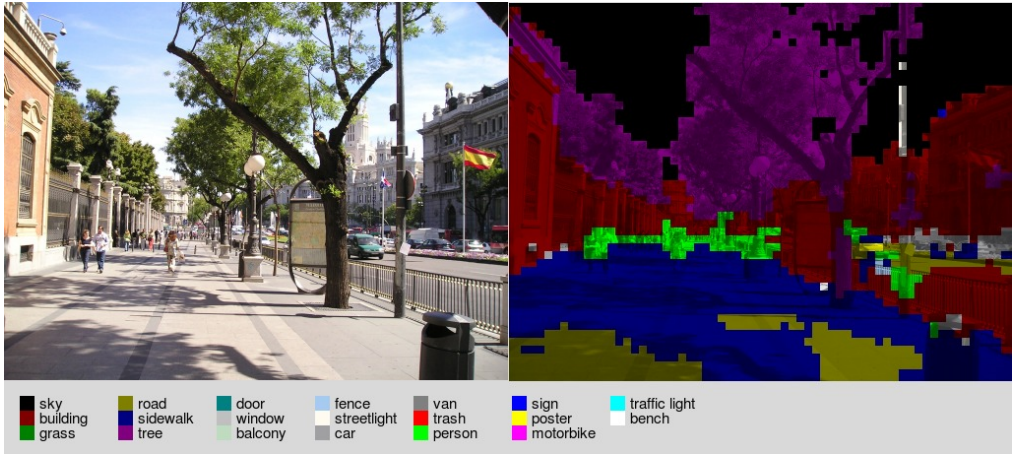


Image: http://david.grangier.info/scene_parsing/

Image Recognition

Main subfields:

› Classification
›› Object detection
››› Semantic segmentation

› **Identification**

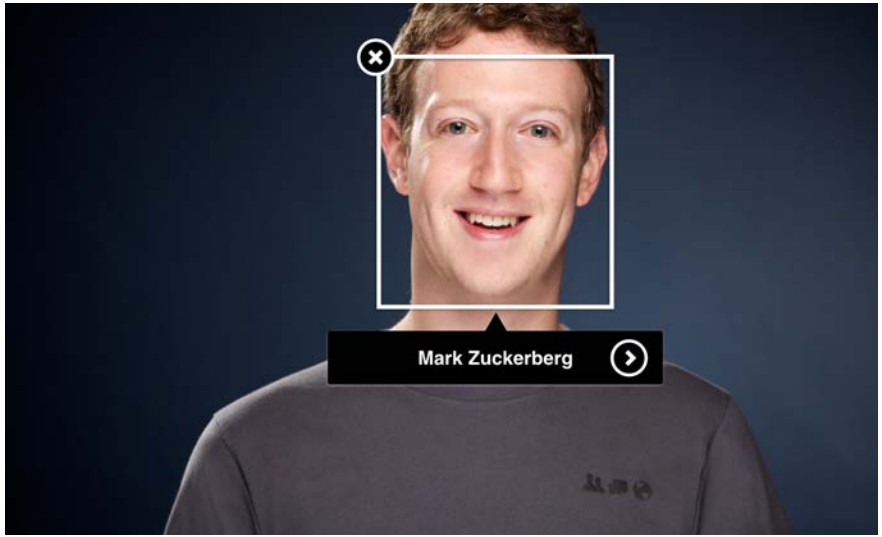
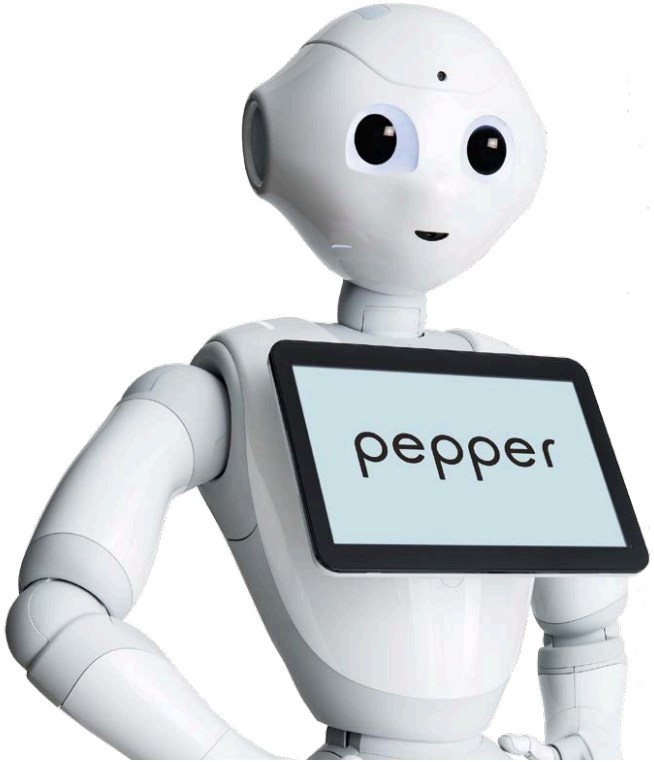


Image: https://bigmedium.com/speaking/design-in-the-era-of-the-algorithm.html

Agenda

- Computer Vision
- Image Recognition
- Pepper
  - Softbank Robotics
  - Characteristics
  - Usecases
  - Computer Vision
  - External services
  - On-device CV with CNNs

Softbank Robotics



- Leader in Humanoid Robotics
- Previously: Aldebaran Robotics
- Headquartered in Paris
- Creator of NAO, Pepper and Romeo
- The robots are used in > 70 countries
- Fields: research, education, retail, healthcare, tourism, hospitality or entertainment, SoftBank

- **Ecosystem of Certified Partners**

SoftBank Robotics

Image: Softbank Robotics

Softbank Robotics

Image: Softbank Robotics

Agenda

- Computer Vision
- Image Recognition
- Pepper
  - Softbank Robotics
  - Characteristics
  - Usecases
  - Computer Vision
  - External services
  - On-device CV with CNNs

Pepper: Technical Characteristics

120 cm
28 kg
20° of freedom
360° movement
Wi-Fi connection
12h of autonomy

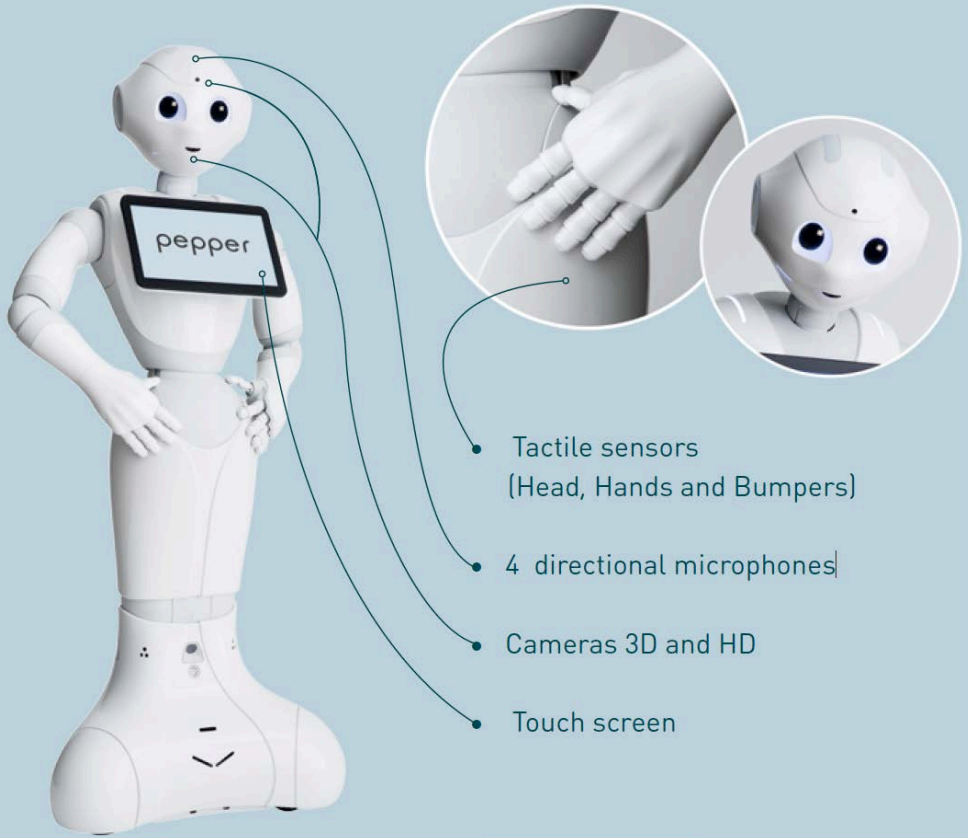


Image: Softbank Robotics

Pepper: Technical Characteristics

Processor	Atom E3845
CPU	Quad core
Clock speed	1.91 GHz
RAM	4 GB DDR3
OS	Nao QI OS
Sensors	2 HD Cameras (OV5640) 1 3D Sensor (ASUS XTION) 4 Microphones 3-axis Gyrometer + 3-axis Accelerometer 6 laser line generators 2 Infra-Red sensors 2 ultrasonic sensors 3 tactile sensors 3 bumpers
Motors	20 Motors and actuators

Tablet

Processor	ARM Cortex-A7
CPU	Quad core
Clock speed	1.3 GHz
RAM	DDR3 SDRAM 1GB (512MB * 2)
Flash Memory	32GB (eMMC)
OS	Android 6.0
Dimensions	246 x 175 x 14.5 mm
Display	Type: IPS, Resolution: 1280*800 Color: 24bit true color
Touch Panel	Capacitive Multi-Touch (5 point)
Sensors	Light illumination, Acceleration Gyro, Geomagnetic

Agenda

- Computer Vision
- Image Recognition
- Pepper
  - Softbank Robotics
  - Characteristics
  - Usecases
  - Computer Vision
  - External services
  - On-device CV with CNNs

So, what can this ‘Pepper’ actually do?

Pepper is:

- Proactive
- Attractive
- Interactive
- Emotional and empathetic
- Connected
- Customisable

So, what can this ‘Pepper’ actually do?

Usecases:

- Welcoming
- Informing and recommending products
- Guiding
- Attracting
- Improving customer knowledge
- Entertaining
- …

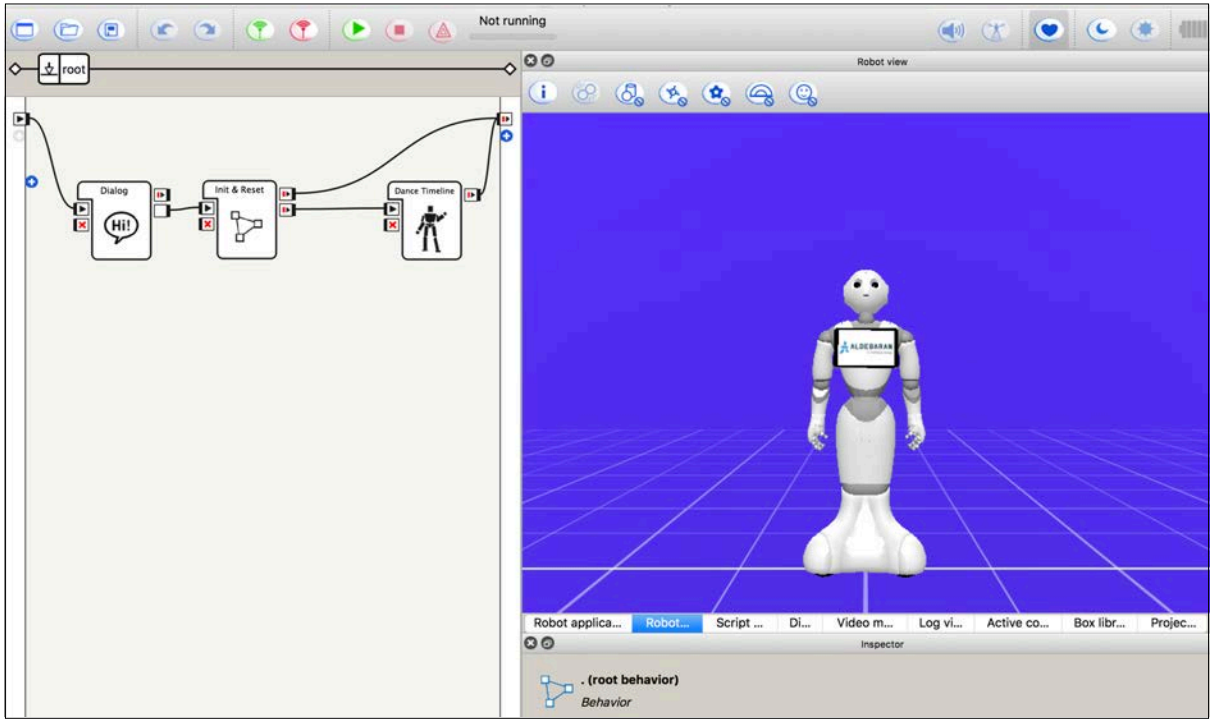
Pepper is still a work in progress…

How to program Pepper

- Choregraphe & Python
- Python SDK
- C++ SDK
- Javascript (Tablet)
- Soon: Android (still reduced function set)
- QiChat (Dialogs)

How to program Pepper

Choregraphe



How to program Pepper
QiChat

```
# Volume
##down
u:([
    "{~can_you} {"ein bisschen" etwas} leiser [sprechen reden]"
    "sprich {"ein bisschen" etwas} leiser"
    "Dreh die Lautstärke runter"
    "sprich nicht so laut"
    "du sprichst zu laut"
])
^gotoReactivate(decrease_volume)
u:($empty) %decrease_volume
^call(ALVolumeSlider.decreaseVolume()) $Demo/back=1
    c1:(false) es tut mir leid, das ist das Minimum
    c1:(true) okay ich spreche jetzt leiser

        u2:([
            nochmal
            mehr
            "noch {"ein bisschen" etwas} mehr"
            "immer noch zu laut"
            ])
            ^gotoReactivate(decrease_volume)
```

Agenda

- Computer Vision
- Image Recognition
- Pepper
  - Softbank Robotics
  - Characteristics
  - Usecases
  - Computer Vision
  - External services
  - On-device CV with CNNs

Computer Vision with Pepper

- Face Detection and People Tracking
- Face Learning and Recognition
- People Characteristics Perception

Computer Vision with Pepper

People Characteristics Perception

```
PeoplePerception/Person/<ID>/AgeProperties
PeoplePerception/Person/<ID>/ExpressionProperties
PeoplePerception/Person/<ID>/GenderProperties
PeoplePerception/Person/<ID>/SmileProperties
PeoplePerception/Person/<ID>/FacialPartsProperties
PeoplePerception/Person/<ID>/Distance
PeoplePerception/Person/<ID>/IsFaceDetected
PeoplePerception/Person/<ID>/IsVisible
PeoplePerception/Person/<ID>/NotSeenSince
PeoplePerception/Person/<ID>/PresentSince
PeoplePerception/Person/<ID>/RealHeight
PeoplePerception/Person/<ID>/ShirtColor
```

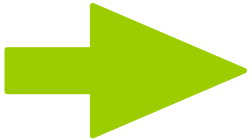
Computer Vision with Pepper

- Face Detection and People Tracking
- Face Learning and Recognition
- People Characteristics Perception
- Emotion Recognition

Computer Vision with Pepper

Emotion Recognition Module

- Data sources:
  - Expression and smile
  - Acoustic voice emotion analysis
  - Head angles
  - Touch sensors
  - Semantic analysis from speech
  - Sound level and energy level of noise
  - Movement detection

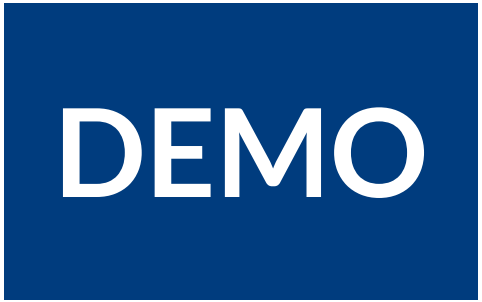


```
Valence
Attention Level
Smile
Expression
{
     “calm"
     “anger"
     “joy"
     "sorrow"
     "laughter"
     "excitement"
     "surprise"
}
(Real values normalized)
```

Computer Vision with Pepper

- People Tracking
- Face Detection, Learning and Recognition
- People Perception
- Emotion Recognition
- Vision Recognition
- Barcode Reader

Computer Vision with Pepper



People Perception

Emotion Recognition

Face Detection and Recognition

Agenda

- Computer Vision
- Image Recognition
- Pepper
  - Softbank Robotics
  - Characteristics
  - Usecases
  - Computer Vision
  - External services
  - On-device CV with CNNs

External services integration

Google's Machine Learning Cloud Vision API

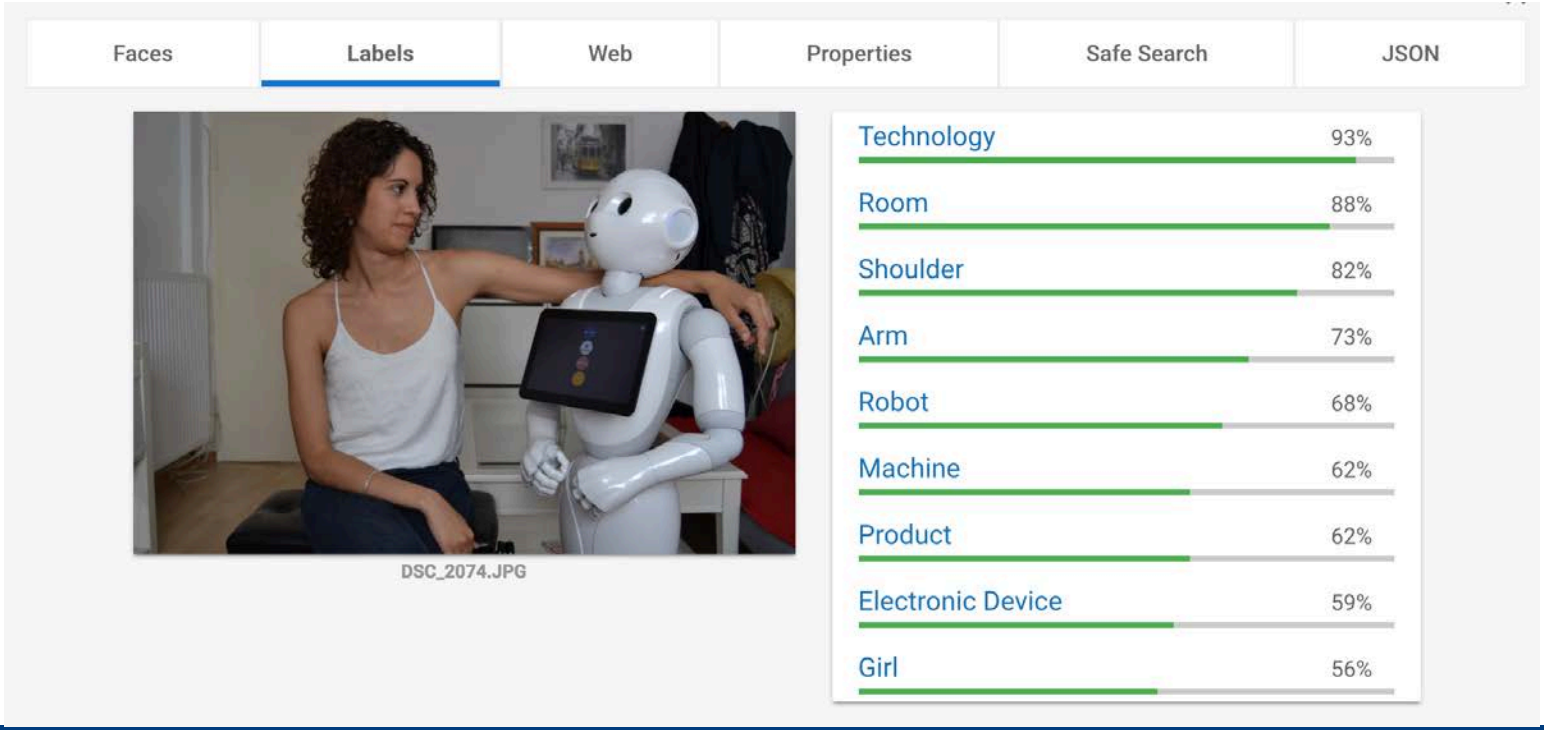
- Machine learning service with pre-trained models
- JSON REST API / client libraries (C#, GO, Java, Node.js, PHP, Python, Ruby)

Explicit Content Detection
Logo Detection
Label Detection
Landmark Detection
Optical Character Recognition
Face Detection
Image Attributes
Web Detection

Google Cloud Platform

External services integration

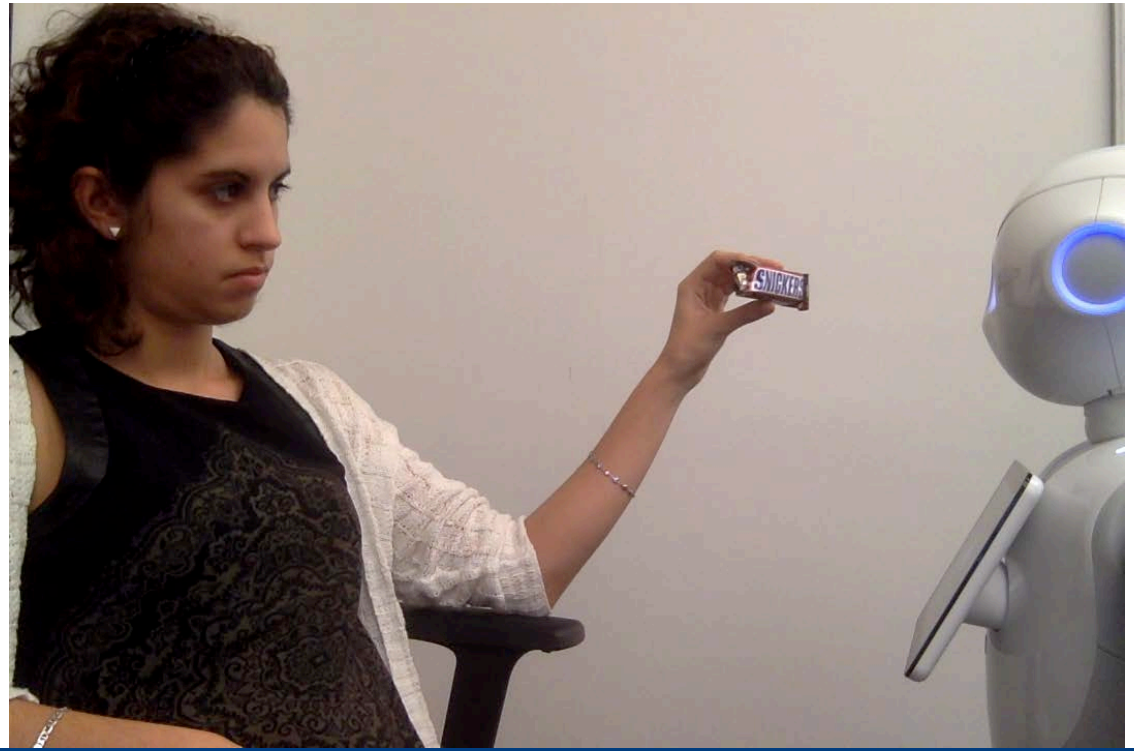
Google's Machine Learning Cloud Vision API: LABELS



https://cloud.google.com/vision/

External services integration

Google's Machine Learning Cloud Vision API: LOGO DETECTION



Good, but show me the code…

External services integration

Google's Machine Learning Cloud Vision API: LABELS

```python
def detect_labels(path):
    """Detects labels in the file."""
    client = vision.ImageAnnotatorClient()

    with io.open(path, 'rb') as image_file:
        content = image_file.read()

    image = types.Image(content=content)

    response = client.label_detection(image=image)
    labels = response.label_annotations
    print('Labels:')

    for label in labels:
        print(label.description)
```

client libraries (C#, GO, Java, Node.js, PHP, **Python**, Ruby)

External services integration

Google's Machine Learning Cloud Vision API: LABELS

```
POST https://vision.googleapis.com/v1/images:annotate?key=YOUR_API_KEY
```

```
{
  "requests": [
    {
      "image": {
        "content": "/9j/7QBEUGhvdG9zaG9...base64-encoded-image-content...fXNWzvDE
      },
      "features": [
        {
          "type": "LABEL_DETECTION"
        }
      ]
    }
  ]
}
```

JSON REST API

External services integration

Google's Machine Learning Cloud Vision API

DEMO

Logo Detection
Label Detection
Optical Character Recognition
Web Detection
Emotion Detection

External services integration

Microsoft Cognitive Services

- Machine learning service with pre-trained models
- JSON REST APIS / client libraries (C#, Android, Swift)

Computer Vision API:
- Analyze Image
- Optical Character Recognition
- Handwritten Text Detection

Face API

Emotion API

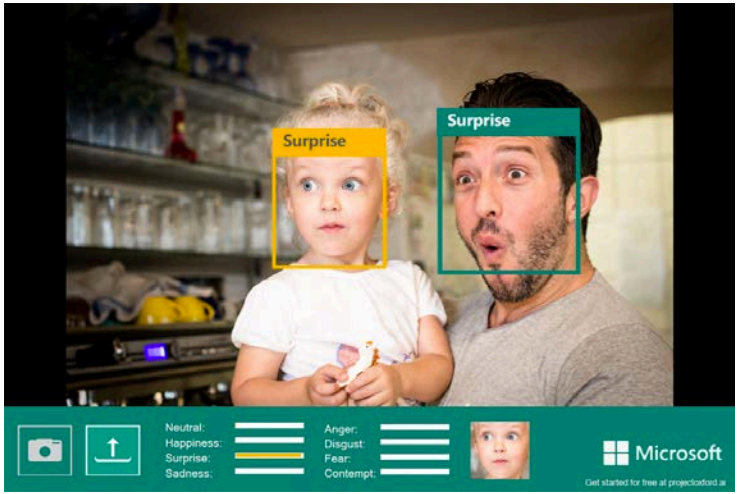
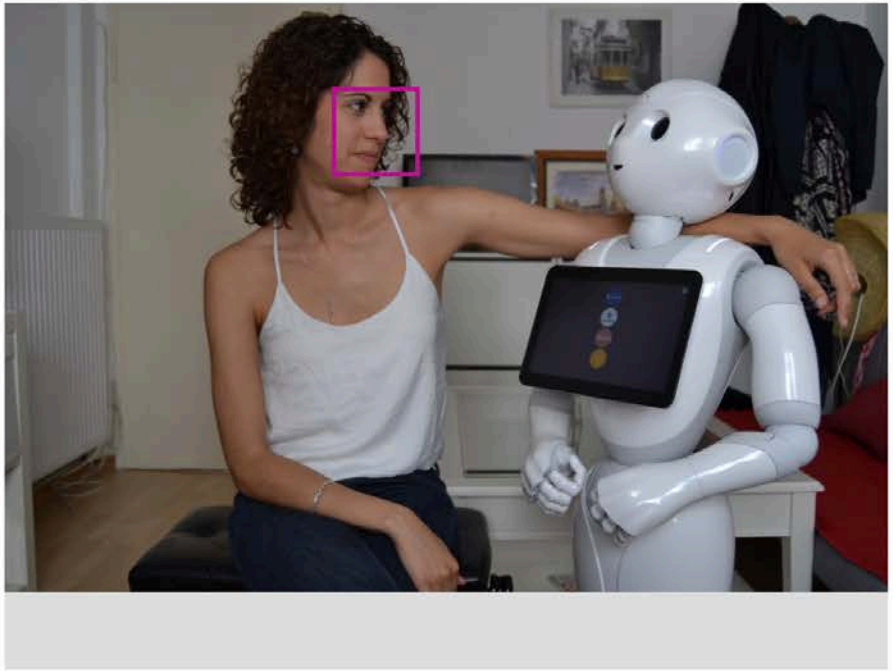


Image: blogs.microsoft.com

External services integration

Microsoft Cognitive Services: COMPUTER VISION API



FEATURE NAME:	VALUE
Description	{ "tags": ["person", "indoor", "woman", "holding", "man", "front", "table", "white", "black", "standing", "young", "cake", "playing", "dog", "room", "plate", "remote", "kitchen"], "captions": [{ "text": "a woman standing in front of a cake", "confidence": 0.75387466 }] }
Tags	[{ "name": "wall", "confidence": 0.996851742 }, { "name": "person", "confidence": 0.996797 }, { "name": "indoor", "confidence": 0.973828435 }]
Image format	"Jpeg"
Image dimensions	1080 x 1620

https://azure.microsoft.com/en-us/services/cognitive-services/computer-vision/

External services integration

Microsoft Cognitive Services

DEMO

Analyze Image

Optical Character Recognition

Handwritten Text Recognition

Emotion Detection

Agenda

- Computer Vision
- Image Recognition
- Pepper
  - Softbank Robotics
  - Characteristics
  - Usecases
  - Computer Vision
  - External services
  - On-device CV with CNNs

On-device CV with CNNs

Why

- Privacy
- Latency
- Connectivity
- Security
- Cost

On-device CV with CNNs

Limitations

- Compute
- Memory
- Storage
- Power
- Bandwidth

On-device CV with CNNs

How

- e.g. Tensorflow Mobile
- or Tensorflow Lite

Pre-trained models

Tensorflow Object Detection API

TensorFlow

Image: softbank.jp

…Out of curiosity

Google's Algorithm found houses

Google's Algorithm found no houses

Yes, but: chihuahua or muffin?

Image: https://medium.freecodecamp.org/chihuahua-or-muffin-my-search-for-the-best-computer-vision-api-cbda4d6b425d

Comparison

- **Amazon's Rekognition** is not just good at identifying the primary object but also the many objects around it
- **Google's Vision API and IBM Watson Vision** return straightforward, descriptive labels
- **Microsoft**'s tags were usually too high level
- **Cloudsight** is a hybrid between human tagging and machine labelling. More accurate. Slower. More expensive.
- **Clarifai** returns, by far, the most tags (at 20) although very generic tags. It also adds qualitative and subjective labels, such as "cute", "funny", "adorable", and "delicious"

https://medium.freecodecamp.org/chihuahua-or-muffin-my-search-for-the-best-computer-vision-api-cbda4d6b425d

Vielen Dank

Silvia Santano
Application Development

Twitter: @SilviaSantano
LinkedIn: silviasantano
ssantano@inovex.de