Introduction to Robotics for cognitive science

Dr. Andrej Lúčny

KAI FMFI UK

lucny@fmph.uniba.sk

Web page of the subject

www.agentspace.org/kv

Why is DL possible today and was not possible before?

Software inventions:

- Dropout & Batch normalization (solves overfitting & missing gradient)
- Xavier initialization
- Novel error functions (metric loss function)

Hardware inventions:

- Big Data Storages
- Graphics Processing Units

Why is DL possible today and was not possible before?

Software inventions:

- Dropout & Batch normalization (solves overfitting & missing gradient)
- Xavier initialization
- Novel error functions (metric loss function)

Hardware inventions:

- Big Data Storages
- Graphics Processing Units

We need a very powerful hardware which is not available everywhere even for use of the DL models

Solution: Cloud technology

- Instead of calling a local subroutine, program compose http request with attached marshaled arguments and get a marshaled result as a response

computer network cloud

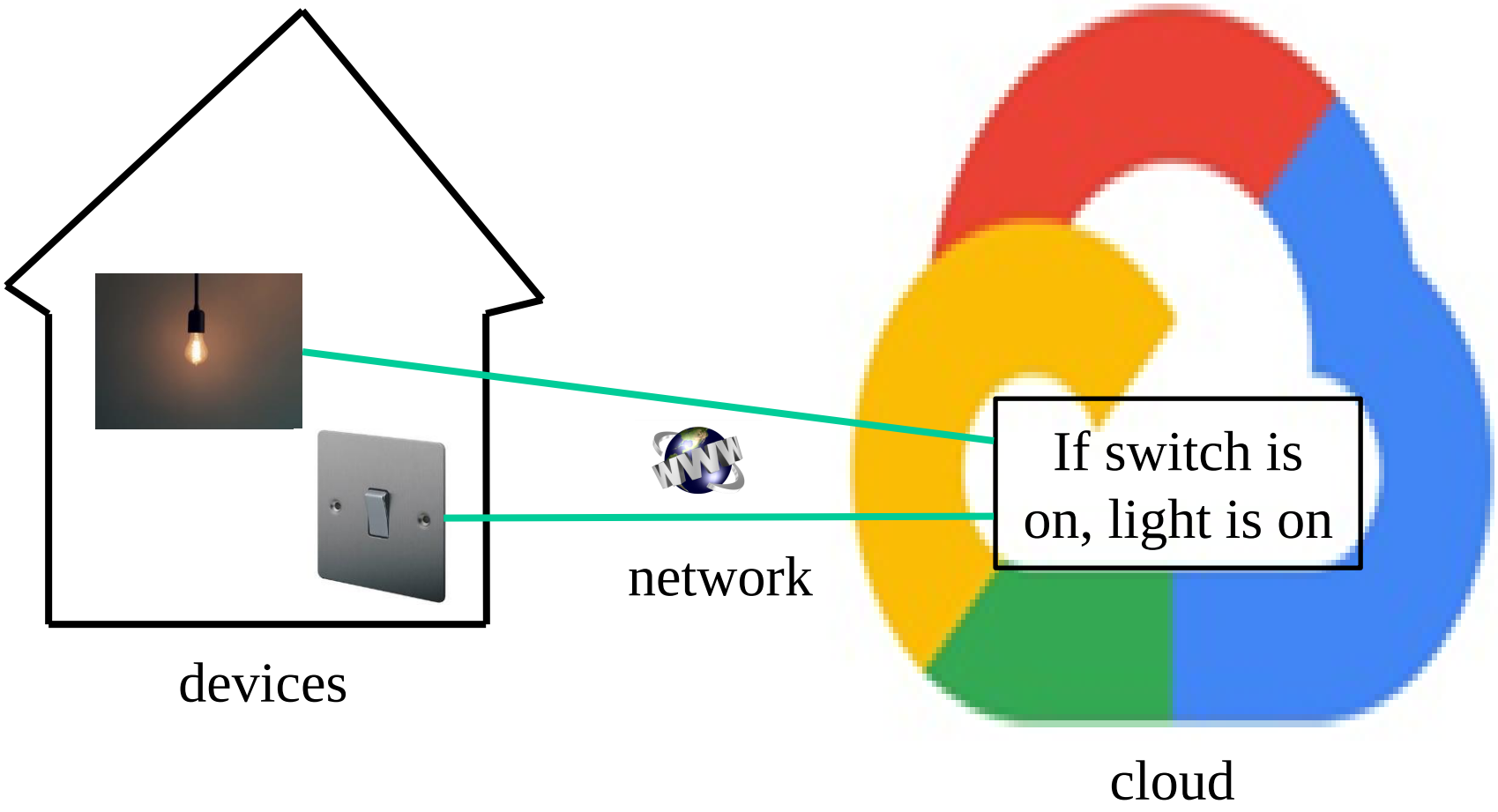
- Today such call takes 80 ms from EU, 40 ms from USA

Calling a cognitive web service

```
import requests
import urllib3
import numpy as np
import cv2

# define the URL to our face detection API
service_url = "http://api.pyimagesearch.com/face_detection/detect/"
image_url = "http://dai.fmph.uniba.sk/upload/1/1a/KUZ2009-3.jpg"

# use our face detection API to find faces in images via image URL
payload = {"url": image_url}
r = requests.post(service_url, data=payload).json()
print(r) # this is result of the web service call
```

{'num_faces': 2, 'success': True, 'faces': [[443, 493, 510, 560], [371, 110, 493, 232]]}

Internet of Things (IoT)



Robot Pepper

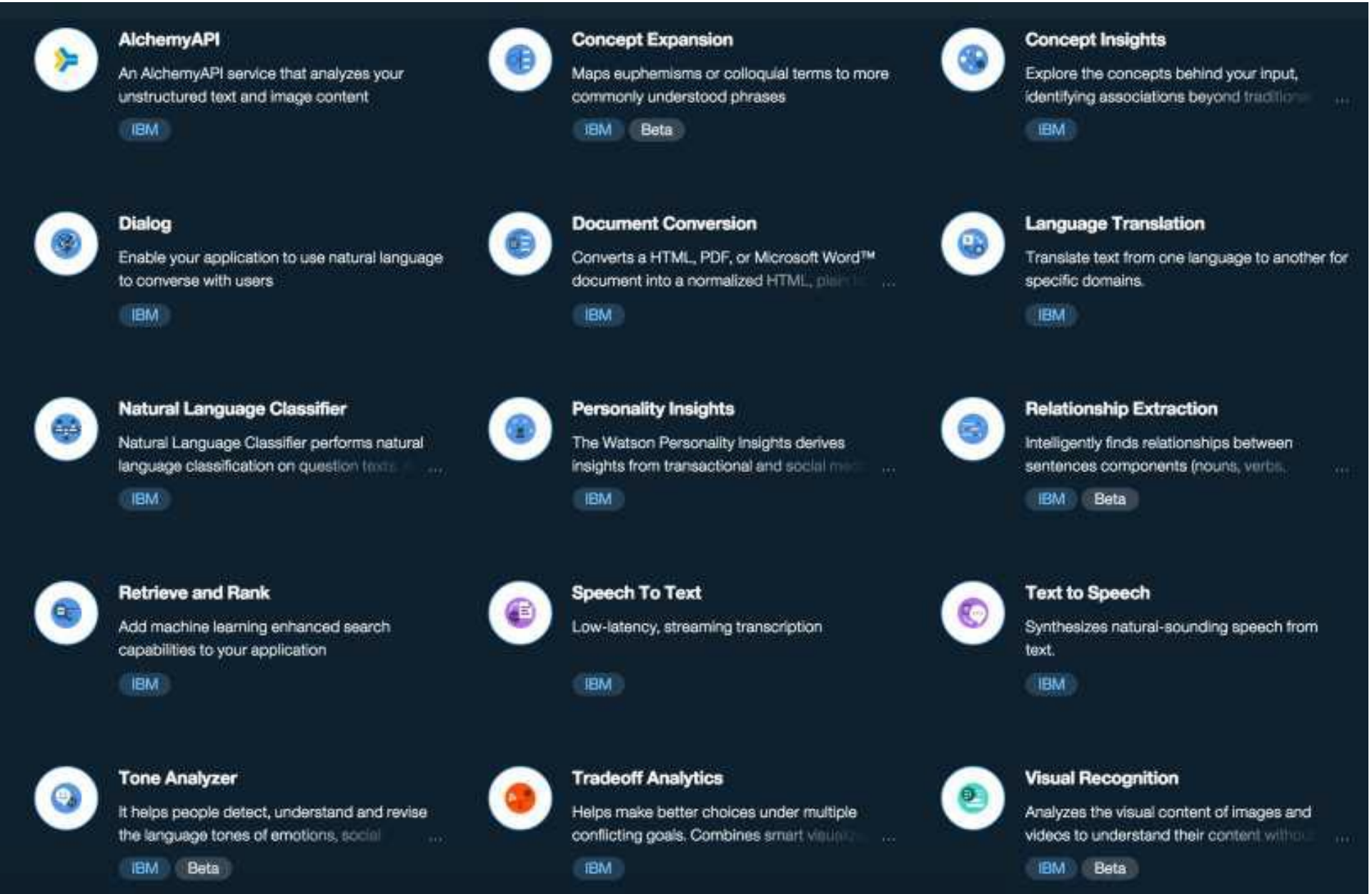
- Relatively cheap robot
- Lower quality
- Calling clould cognitive services, e.g. face recognition
- Without connection to the Internet is not working

Google Cloud

Google cloud provides APIs for computer vision, speech recognition, natural language processing, and translation.

- *Google Cloud Video Intelligence API* makes videos searchable and discoverable by extracting metadata, identifying key nouns, and annotating the content of the video.
- *Google Cloud Vision API* enables you to understand the content of an image including categories, objects and faces, words, and more. Face recognition is a common use of Vision API.
- *Google Cloud Speech API* enables you to convert audio to text by applying neural network models in an easy to use API.
- *Google Natural Language API* provides developers functionality to information about people, places, events and much more, mentioned in text documents, news articles or blog posts.
- *Google Cloud Translation API* lets developers convert text from a source language to a target language.

IBM Watson


AlchemyAPI
An AlchemyAPI service that analyzes your unstructured text and image content
IBM
Concept Expansion
Maps euphemisms or colloquial terms to more commonly understood phrases
IBM
Beta
Concept Insights
Explore the concepts behind your input, identifying associations beyond
IBM
Dialog
Enable your application to use natural language to converse with users
IBM
Document Conversion
Converts a HTML, PDF, or Microsoft Word™ document into a normalized HTML,
IBM
Language Translation
Translate text from one language to another for specific domains.
IBM
Natural Language Classifier
Natural Language Classifier performs natural language classification on question
IBM
Personality Insights
The Watson Personality Insights derives insights from transactional and social
IBM
Relationship Extraction
Intelligently finds relationships between sentences components (nouns, verbs,
IBM
Beta
Retrieve and Rank
Add machine learning enhanced search capabilities to your application
IBM
Speech To Text
Low-latency, streaming transcription
IBM
Text to Speech
Synthesizes natural-sounding speech from text.
IBM
Tone Analyzer
It helps people detect, understand and revise the language tones of emotions, social
IBM
Beta
Tradeoff Analytics
Helps make better choices under multiple conflicting goals. Combines smart visual
IBM
Visual Recognition
Analyzes the visual content of images and videos to understand their content
IBM
Beta

MicroSoft Azure

Cognitive Services APIs

Vision API
Computer Vision
Custom Vision Service
Face API
Forms Recognizer PREVIEW
Ink Recognizer PREVIEW
Video Indexer

Speech API
Speech Services
Speaker Recognition PREVIEW
Bing Speech API RETIRING
Translator Speech RETIRING

Search API
Bing News Search
Bing Video Search
Bing Web Search
Bing Autosuggest
Bing Custom Search
Bing Entity Search
Bing Image Search
Bing Visual Search
Bing Spell Check
Bing Local Business Search PREVIEW

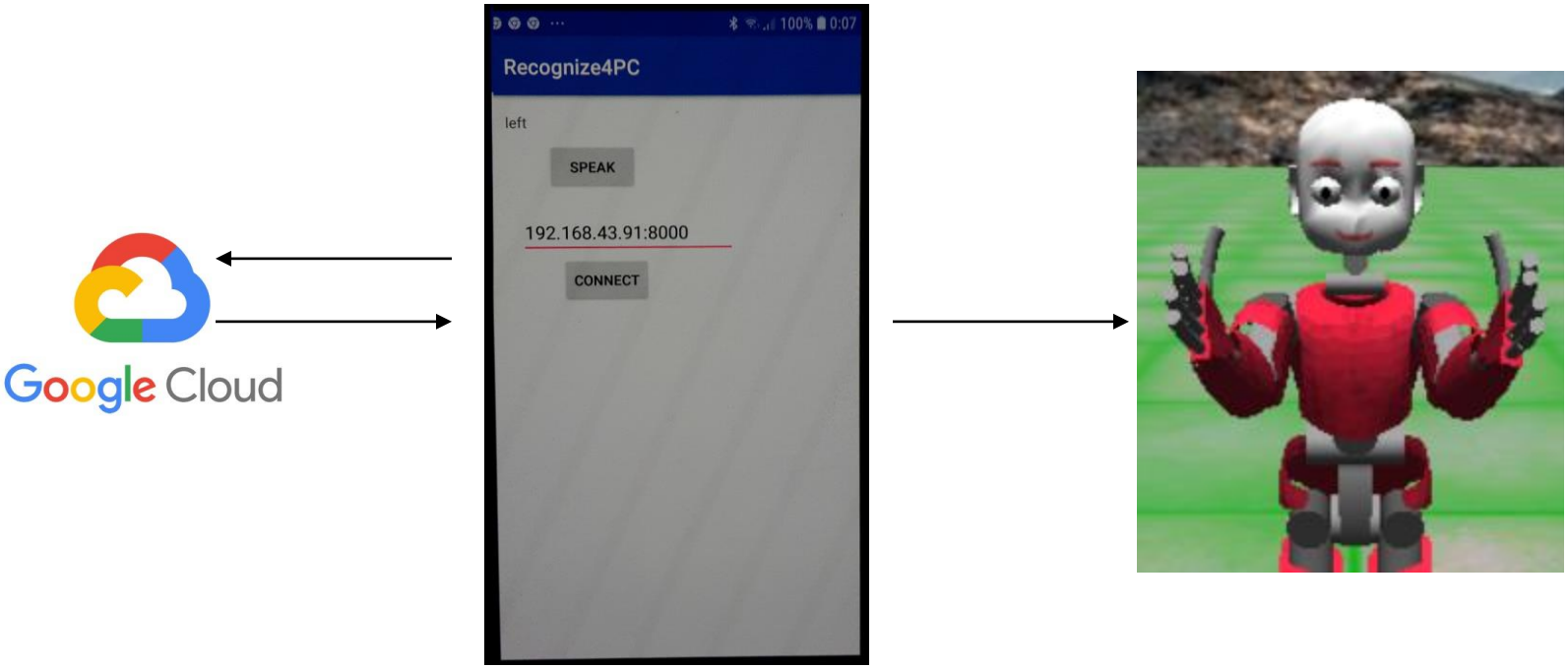
Decision API
Anomaly Detector PREVIEW
Content Moderator
Personalizer PREVIEW

Language API
Language Understanding (LUIS)
QnA Maker
Text Analytics
Translator Text

A dark side of the cloud services

- Cloud services could be very comfortable
- However they are not -
- - because of their business model
- - each user must register
- - each call is charged
- - quality can be disputable and service rather freely collects data from users
- One can get free period or free initial amount of calls
- Exception: Android platform can call Google cloud without any restrictions

Voice recognition from Android

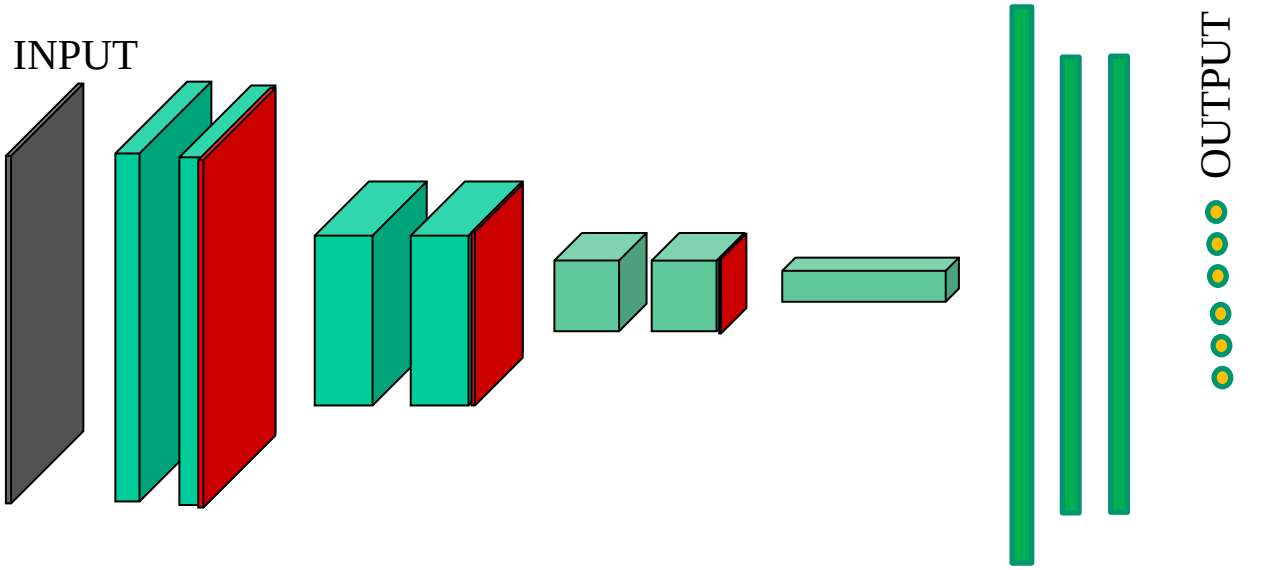


https://github.com/andylucny/Recognize4PC

```
Intent intent = new Intent(RecognizerIntent.ACTION_RECOGNIZE_SPEECH);
// Specify the calling package to identify your application
intent.putExtra(RecognizerIntent.EXTRA_CALLING_PACKAGE,
    getClass().getPackage().getName());

// Given an hint to the recognizer about what the user is going to say
intent.putExtra(RecognizerIntent.EXTRA_LANGUAGE_MODEL,
    RecognizerIntent.LANGUAGE_MODEL_WEB_SEARCH);

int noOfMatches = 1;
// Specify how many results you want to receive. The results will be
// sorted where the first result is the one with higher confidence.
intent.putExtra(RecognizerIntent.EXTRA_MAX_RESULTS, noOfMatches);

startActivityForResult(intent, VOICE_RECOGNITION_REQUEST_CODE);

protected void onActivityResult(int requestCode, int resultCode, Intent data) {
    if (requestCode == VOICE_RECOGNITION_REQUEST_CODE) {
        //If Voice recognition is successful then it returns RESULT_OK
        if (resultCode == RESULT_OK) {
            ArrayList<String> textMatchList =
                data.getStringArrayListExtra(RecognizerIntent.EXTRA_RESULTS);
            if (!textMatchList.isEmpty()) {
                // populate the Matches
                sendRecognizedText(textMatchList.get(0));
            }
}}}
```

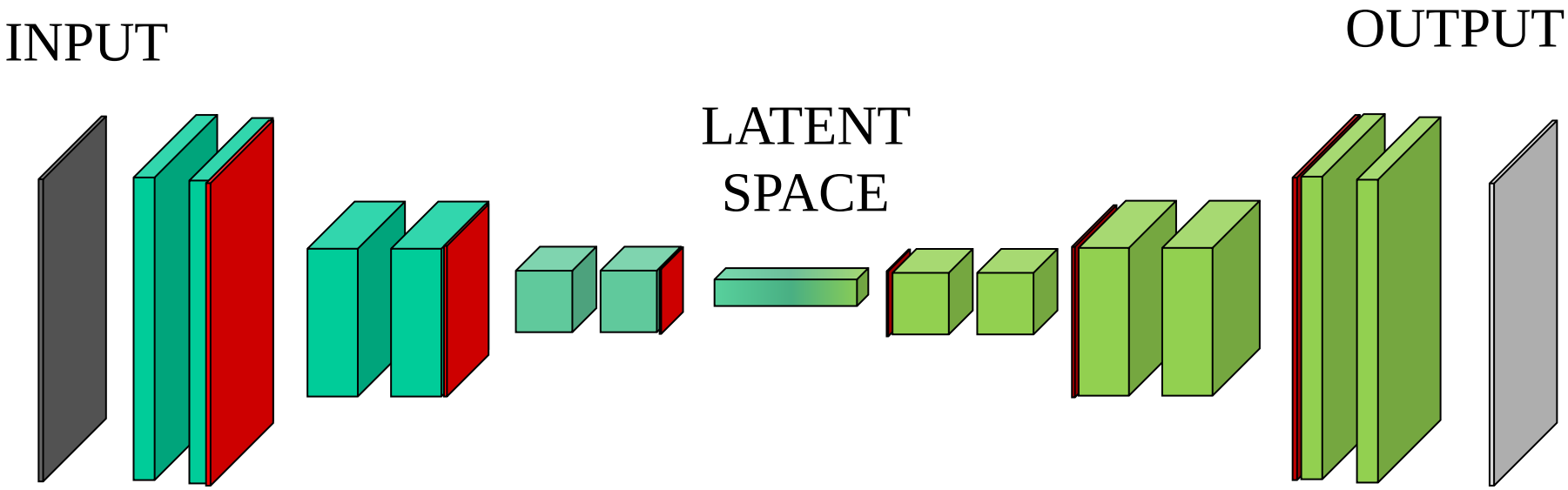
How the cloud cognitive services works?

Classifiers and detectors

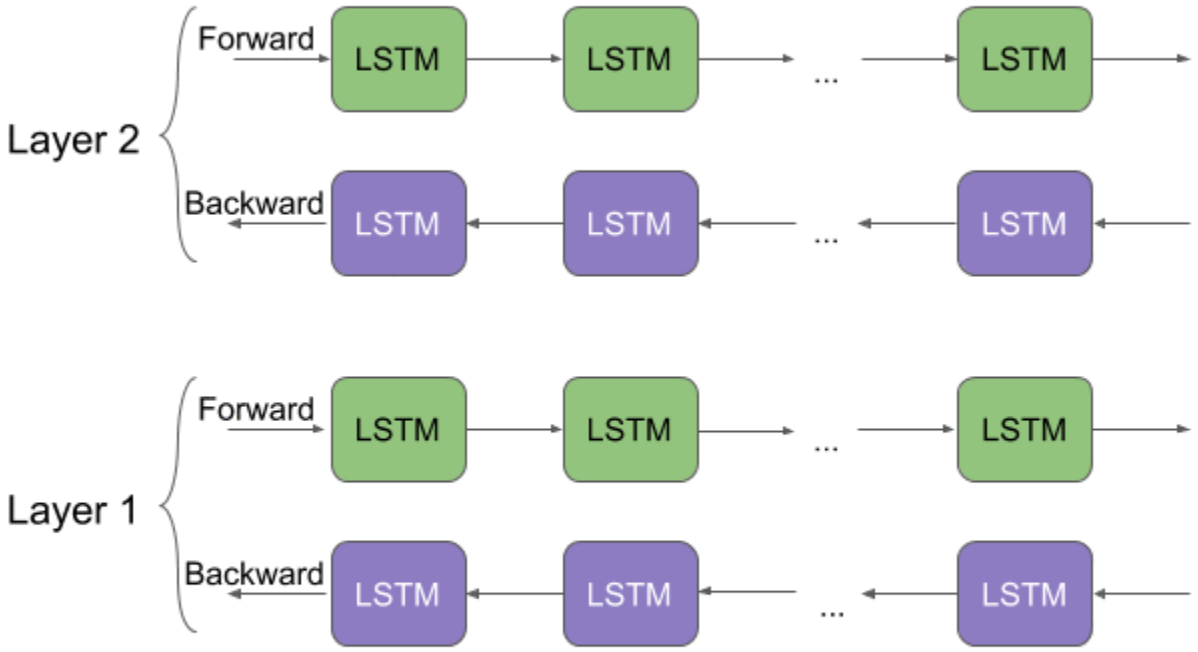


Captioning, Translators

Encoder - Decoder



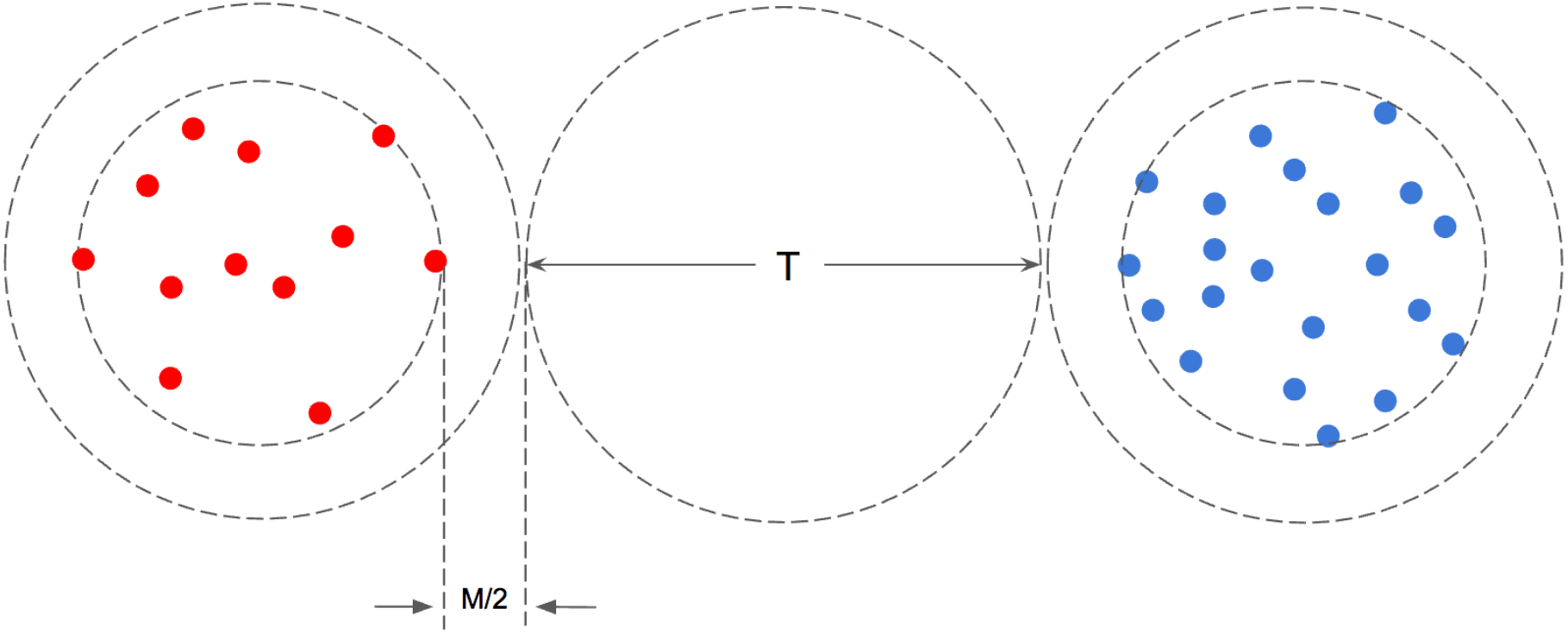
Processing of speech, text, video



Deep Recurrent Neural Networks

Metric Loss Function

- Samples contains just category though we look for a value
- Good results: the same category & distance < T-M or different category & distance > T+M, Bad results: otherwise
- gradient is estimated from a few worst cases, e.g. red dot close to blue dot should be moved from the blue dot



Vectorization for Recognition

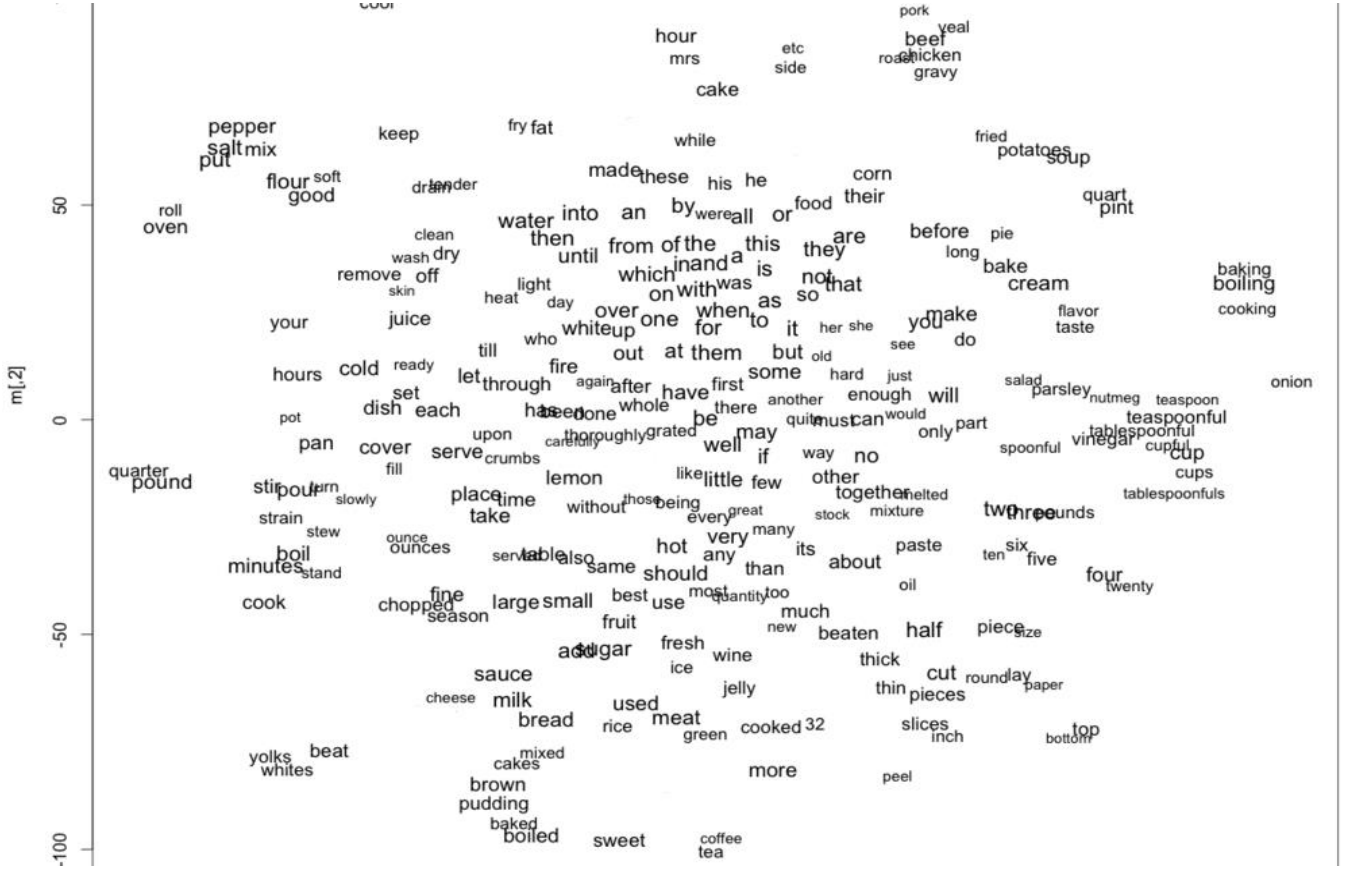
RESNET (0.453, 0.122, 0.998, …) ENROLL Lucny

Ordinary ResNet network trained by the metric loss function on dataset containing several faces for each person can provide vector of 256 floats <0,1> such that:

- vectors provided for faces of the same person are similar (their Euclid distance is low)
- vectors provided for faces of different persons are different (their Euclid distance is high)

`http://www.agentspace.org/andy/learnopencv/`

New generation of chatbots



- Based on vectorization of words and sentences
- For each intention one enter few examples how we can ask
- Vector of one example is closest to given question