Adding Intelligence to Your Application

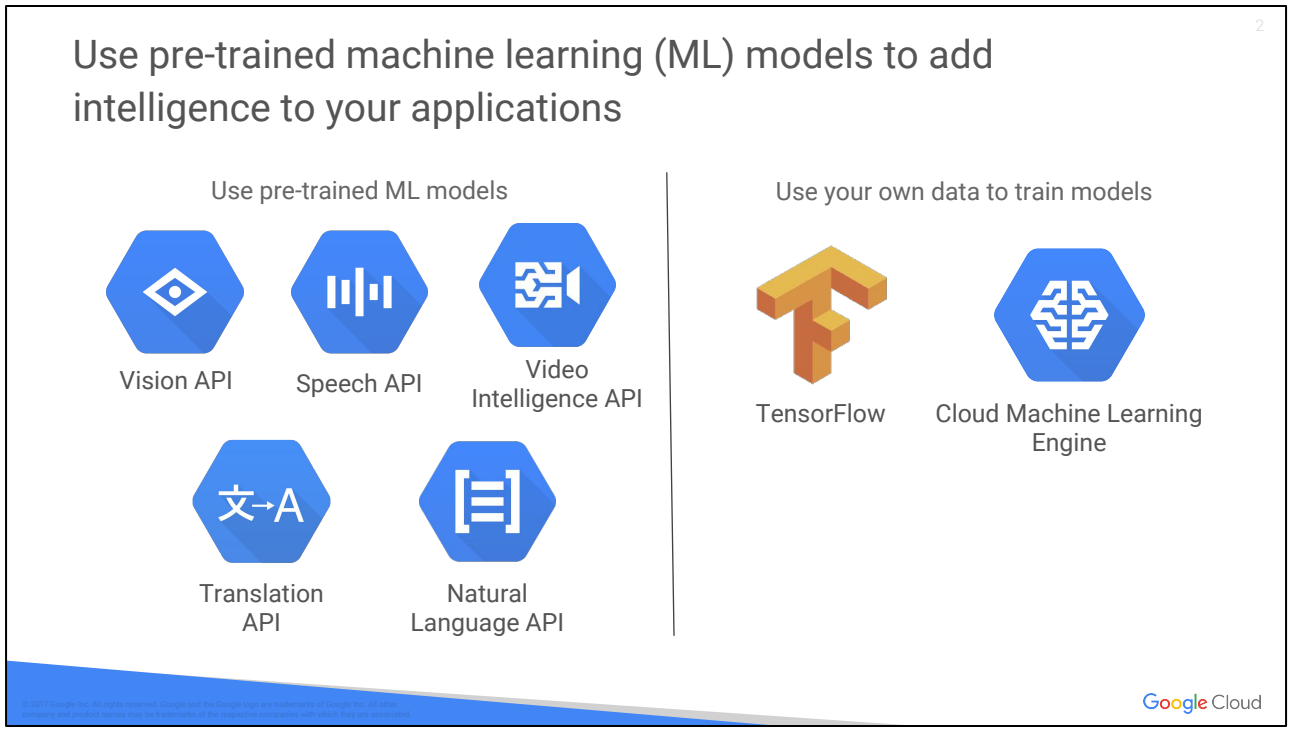
Developing Applications with Google Cloud Platform

Vision API, Speech API, Video Intelligence API, Translation API, Natural Language API

Version 2.0
Last modified: 2017-09-18

Google Cloud

This presentation gives you a brief overview of the pre-trained machine learning APIs in Google Cloud Platform.



Google Cloud Platform offers several pre-trained machine-learning (ML) models that you can use to add intelligence to your application.

- Cloud Vision API enables you to perform complex image detection. For more information, see https://cloud.google.com/vision/.
- Cloud Speech API enables developers to convert audio to text. It handles 110 languages and variants to support your global user base. You can transcribe the text of users dictating to an application's microphone, enable command-and-control through voice, transcribe audio files, and more. For more information, see https://cloud.google.com/speech/.
- Cloud Video Intelligence API enables you to search every moment of every video file to extract and understand the video's entities at the shot, frame, or video level. The API annotates videos stored in Google Cloud Storage and helps you identify key noun entities of your video and when they occur within the video. For more information, see https://cloud.google.com/video-intelligence/.
- Cloud Translation API enables you to translate an arbitrary string into any supported language. Translation API is highly responsive. Websites and applications can use Translation API for fast, dynamic translation of text from a source language to a target language (e.g., Japanese to English). For more information, see https://cloud.google.com/translate/.
- Cloud Natural Language API enables you to extract information about entities such as people, places, and events that are mentioned in text documents,

- news articles, or blog posts. You can use the API to understand sentiment about your product on social media or parse intent from customer conversations. For more information, see https://cloud.google.com/natural-language/.

You can also use your own data to build and train your own ML models by using TensorFlow and Cloud Machine Learning Engine.

Invoke REST APIs to use machine learning APIs; no machine learning knowledge is required

Image (GCS/embedded)

JSON request

Invoke Vision API

The Vision API can work off an image in Cloud Storage or embedded directly into a POST message. I'll use C

. That photograph is from http://www.publicdomainpictures.net/view-image.php?image=15842

```
# Running Vision API
import base64
IMAGE="gs://cloud-training-demos/vision/sign2.jpg"
vservice = build('vision', 'v1', developerKey=APIKEY)
request = vservice.images().annotate(body={
        'requests': [{
                'image': {
                    'source': {
                        'gcs_image_uri': IMAGE
                    }
                },
                'features': [{
                    'type': 'TEXT_DETECTION',
                    'maxResults': 3,
                }]
            }],
        })
responses = request.execute(num_retries=3)
print responses
```

JSON response

{u'responses': [{u'textAnnotations': [{u'locale': u'zh', u'description': u'\u8bf7\u6
\u4f18\n\u7f8e\u6c34\u73af\u5883\n', u'boundingPoly': {u'vertices': [{u'y': 103, u'x
 {u'y': 654, u'x': 150}]}}, {u'description': u'\u8bf7', u'boundingPoly': {u'vertices
41, u'x': 322}, {u'y': 241, u'x': 178}]}}, {u'description': u'\u60a8', u'boundingPol

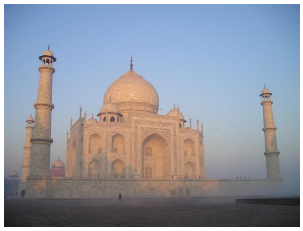
Google Cloud

Examples

Analyze images

Label Detection

Optical character recognition (OCR)



Landmark Detection

Logo Detection

Face Detection

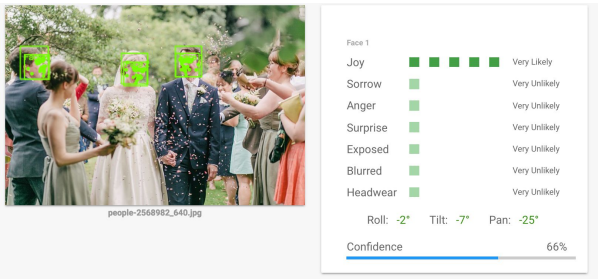
Explicit Content Detection

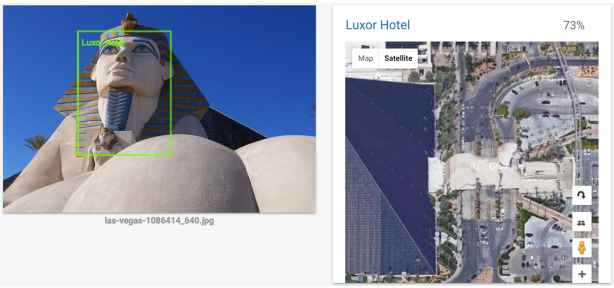
Google Cloud

The Vision API can categorize objects under labels and perform optical character recognition (OCR). The Vision API can detect landmarks, logos, faces, and explicit content.

Images:
https://pixabay.com/en/lion-big-cat-predator-safari-515028/
https://pixabay.com/en/maintenance-under-construction-2422173/
https://pixabay.com/en/taj-mahal-india-agra-tomb-grave-366/
https://pixabay.com/en/google-wood-wooden-brown-620522/
https://pixabay.com/en/smartphone-hand-photo-montage-faces-1445489/
https://pixabay.com/en/soap-bubbles-colorful-fly-2405969/

Get insight from images



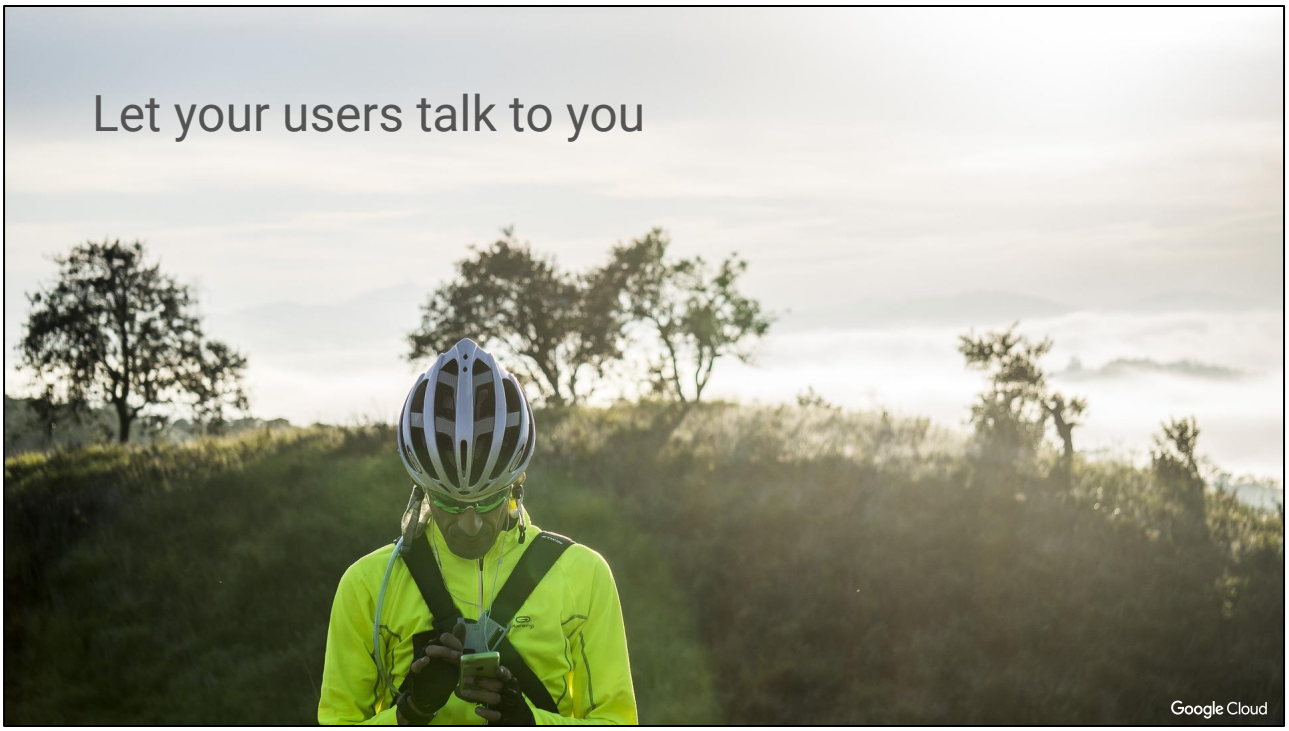


Google Cloud

For example, Vision API can analyze faces and return information about emotions and headwear. In the wedding picture, the API accurately returns the emotional expressions on the faces in the picture.

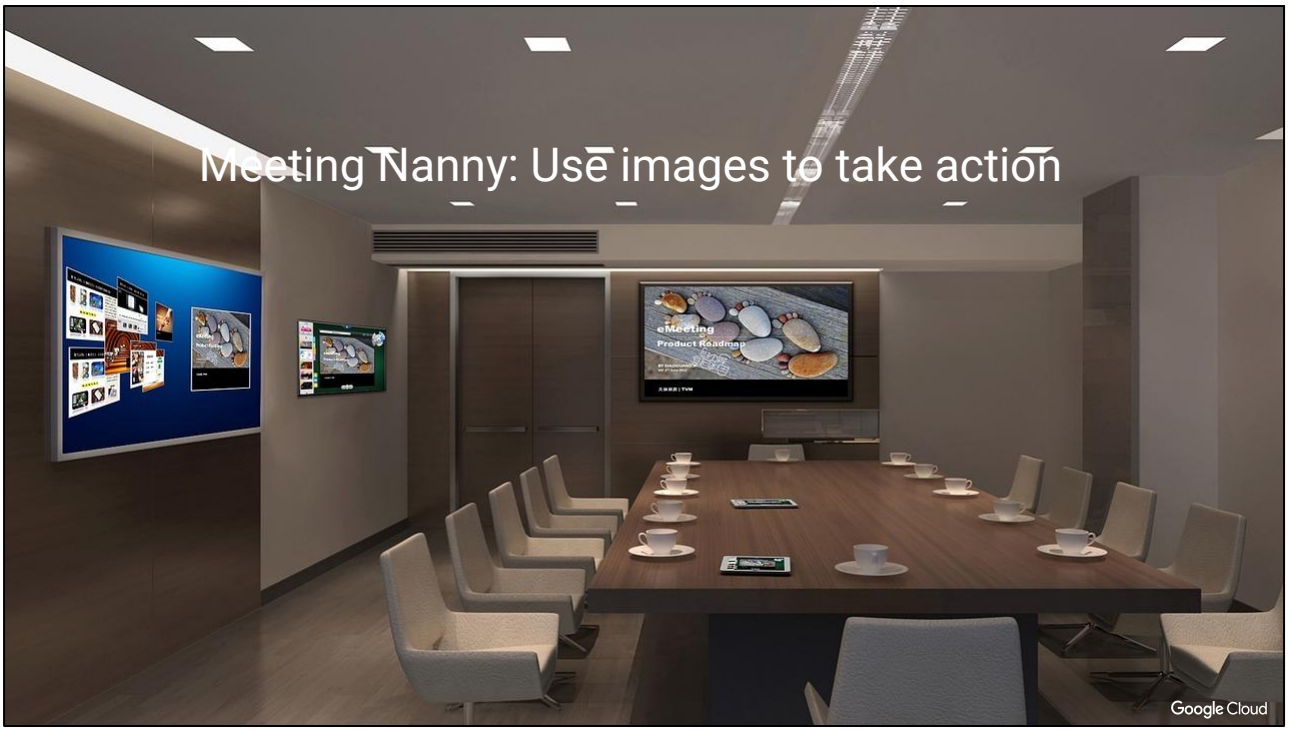
In the picture of the Sphinx, Vision API correctly detects that the image is from the Sphinx in Las Vegas and not the Sphinx in Egypt.

Images:
https://pixabay.com/en/people-crowd-couple-wedding-bride-2568982/
https://pixabay.com/en/las-vegas-pharaoh-egypt-vegas-1086414/



Speech API enables developers to convert audio to text. It handles 110 languages and variants to support your global user base. You can transcribe the text of users dictating to an application's microphone, enable command-and-control through voice, transcribe audio files, and more.

Images:
https://pixabay.com/en/cycling-bike-trail-sport-sol-1533268/



Google’s conference room systems perform occupancy detection by using motion detection with the VC camera and by call ID matching. Every 30 seconds, the VC unit sends a Cloud Pub/Sub notification indicating whether motion was detected or not. It also sends a Cloud Pub/Sub notification when a call starts or ends.
If motion is detected between 6 and 8 minutes after the meeting start time, the room counts as occupied. Otherwise, it's empty.

Images:
https://pixabay.com/en/interior-design-tv-multi-screen-828545/