REAL-LIFE APPLICATION OF SERVERLESS ML

FOR DEVELOPERS

Who we are?

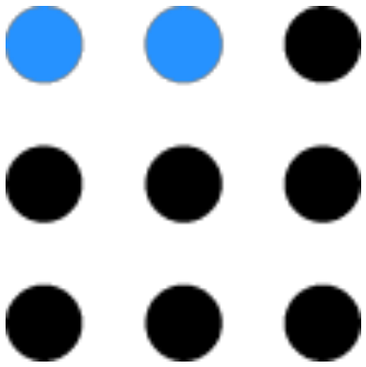
Antonio Piraino

Backend DevOps
@ Quokky

Ali Gholami

CTO & Backend developer
@ Quokky

What are we doing?



https://quokky.com

Cloud Base Document Sharing Platforms
with Web/Mobile/Desktop Applications

https://duedot.co

ML
as a Service

Agenda

Define domain of the problem

Analyse first solution

Machine Learning

How we solved our problem using ML

Going Serverless

Reimplementing the whole service using serverless

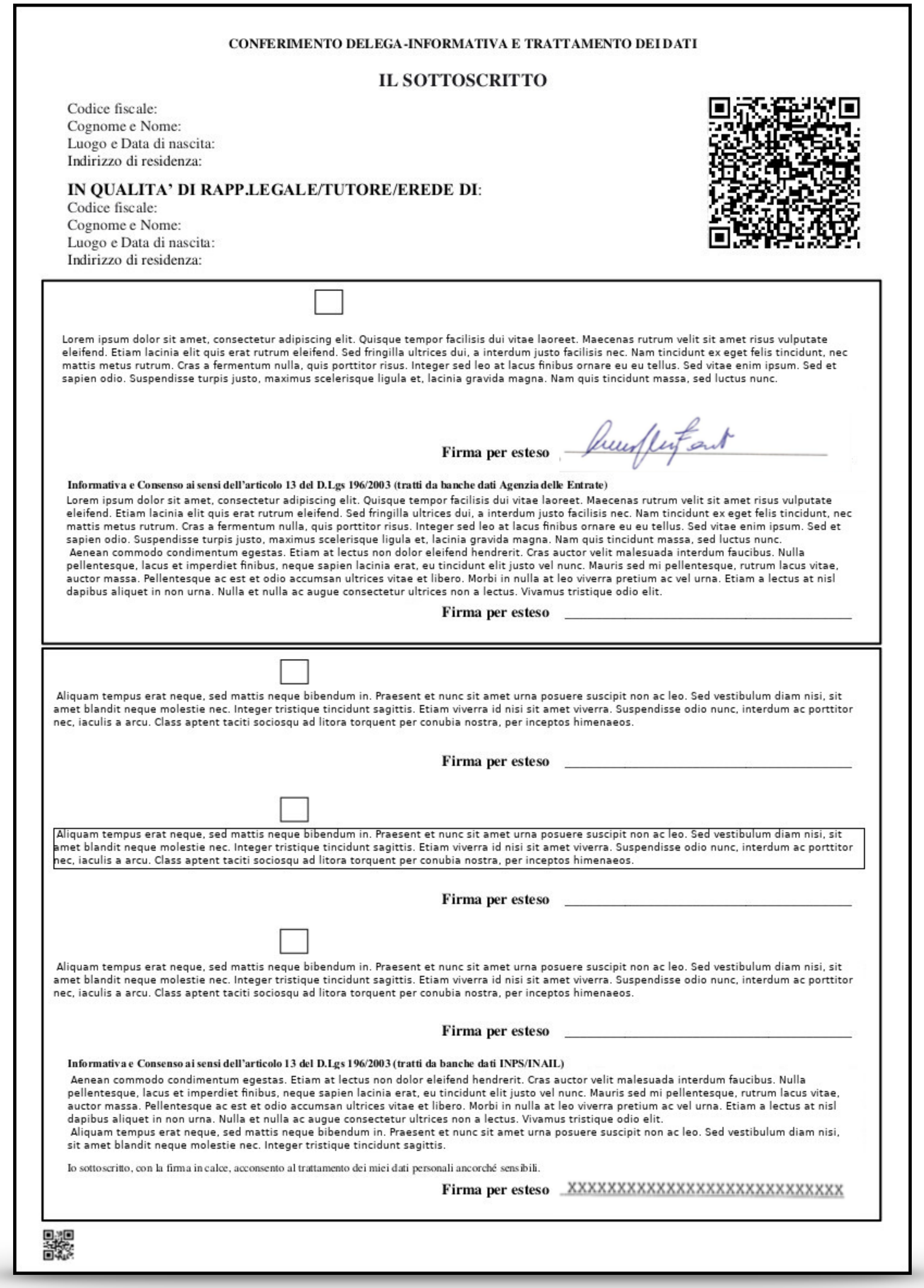
Signature Recognition

Signature recognition system to classify:

Signature 

Placeholder XXXXXXXXXXXXXXXXXXXXXXXXXXXXX

No Signature



CONFERIMENTO DELEGA-INFORMATIVA E TRATTAMENTO DEI DATI

IL SOTTOSCRITTO

Codice fiscale:
Cognome e Nome:
Luogo e Data di nascita:
Indirizzo di residenza:

IN QUALITA' DI RAPP.LEGALE/TUTORE/EREDE DI:
Codice fiscale:
Cognome e Nome:
Luogo e Data di nascita:
Indirizzo di residenza:

☐

Lorem ipsum dolor sit amet, consectetur adipiscing elit. Quisque tempor facilisis dui vitae laoreet. Maecenas rutrum velit sit amet risus vulputate eleifend. Etiam lacinia elit quis erat rutrum eleifend. Sed fringilla ultrices dui, a interdum justo facilisis nec. Nam tincidunt ex eget felis tincidunt, nec mattis metus rutrum. Cras a fermentum nulla, quis porttitor risus. Integer sed leo at lacus finibus ornare eu eu tellus. Sed vitae enim ipsum. Sed et sapien odio. Suspendisse turpis justo, maximus scelerisque ligula et, lacinia gravida magna. Nam quis tincidunt massa, sed luctus nunc.

Firma per esteso ____________________

Informativa e Consenso ai sensi dell'articolo 13 del D.Lgs 196/2003 (tratti da banche dati Agenzia delle Entrate)

Lorem ipsum dolor sit amet, consectetur adipiscing elit. Quisque tempor facilisis dui vitae laoreet. Maecenas rutrum velit sit amet risus vulputate eleifend. Etiam lacinia elit quis erat rutrum eleifend. Sed fringilla ultrices dui, a interdum justo facilisis nec. Nam tincidunt ex eget felis tincidunt, nec mattis metus rutrum. Cras a fermentum nulla, quis porttitor risus. Integer sed leo at lacus finibus ornare eu eu tellus. Sed vitae enim ipsum. Sed et sapien odio. Suspendisse turpis justo, maximus scelerisque ligula et, lacinia gravida magna. Nam quis tincidunt massa, sed luctus nunc.

Aenean commodo condimentum egestas. Etiam at lectus non dolor eleifend hendrerit. Cras auctor velit malesuada interdum faucibus. Nulla pellentesque, lacus et imperdiet finibus, neque sapien lacinia erat, eu tincidunt elit justo vel nunc. Mauris sed mi pellentesque, rutrum lacus vitae, auctor massa. Pellentesque ac est et odio accumsan ultrices vitae et libero. Morbi in nulla at leo viverra pretium ac vel urna. Etiam a lectus at nisl dapibus aliquet in non urna. Nulla et nulla ac augue consectetur ultrices non a lectus. Vivamus tristique odio elit.

Firma per esteso ____________________

☐

Aliquam tempus erat neque, sed mattis neque bibendum in. Praesent et nunc sit amet urna posuere suscipit non ac leo. Sed vestibulum diam nisi, sit amet blandit neque molestie nec. Integer tristique tincidunt sagittis. Etiam viverra id nisi sit amet viverra. Suspendisse odio nunc, interdum ac porttitor nec, iaculis a arcu. Class aptent taciti sociosqu ad litora torquent per conubia nostra, per inceptos himenaeos.

Firma per esteso ____________________

☐

Aliquam tempus erat neque, sed mattis neque bibendum in. Praesent et nunc sit amet urna posuere suscipit non ac leo. Sed vestibulum diam nisi, sit amet blandit neque molestie nec. Integer tristique tincidunt sagittis. Etiam viverra id nisi sit amet viverra. Suspendisse odio nunc, interdum ac porttitor nec, iaculis a arcu. Class aptent taciti sociosqu ad litora torquent per conubia nostra, per inceptos himenaeos.

Firma per esteso ____________________

☐

Aliquam tempus erat neque, sed mattis neque bibendum in. Praesent et nunc sit amet urna posuere suscipit non ac leo. Sed vestibulum diam nisi, sit amet blandit neque molestie nec. Integer tristique tincidunt sagittis. Etiam viverra id nisi sit amet viverra. Suspendisse odio nunc, interdum ac porttitor nec, iaculis a arcu. Class aptent taciti sociosqu ad litora torquent per conubia nostra, per inceptos himenaeos.

Firma per esteso ____________________

Informativa e Consenso ai sensi dell'articolo 13 del D.Lgs 196/2003 (tratti da banche dati INPS/INAIL)

Aenean commodo condimentum egestas. Etiam at lectus non dolor eleifend hendrerit. Cras auctor velit malesuada interdum faucibus. Nulla pellentesque, lacus et imperdiet finibus, neque sapien lacinia erat, eu tincidunt elit justo vel nunc. Mauris sed mi pellentesque, rutrum lacus vitae, auctor massa. Pellentesque ac est et odio accumsan ultrices vitae et libero. Morbi in nulla at leo viverra pretium ac vel urna. Etiam a lectus at nisl dapibus aliquet in non urna. Nulla et nulla ac augue consectetur ultrices non a lectus. Vivamus tristique odio elit.

Aliquam tempus erat neque, sed mattis neque bibendum in. Praesent et nunc sit amet urna posuere suscipit non ac leo. Sed vestibulum diam nisi, sit amet blandit neque molestie nec. Integer tristique tincidunt sagittis.

Io sottoscritto, con la firma in calce, acconsento al trattamento dei miei dati personali ancorché sensibili.

Firma per esteso XXXXXXXXXXXXXXXXXXXXXXXXXXXXX

Quokky

Document Classification

Categorise tax documents

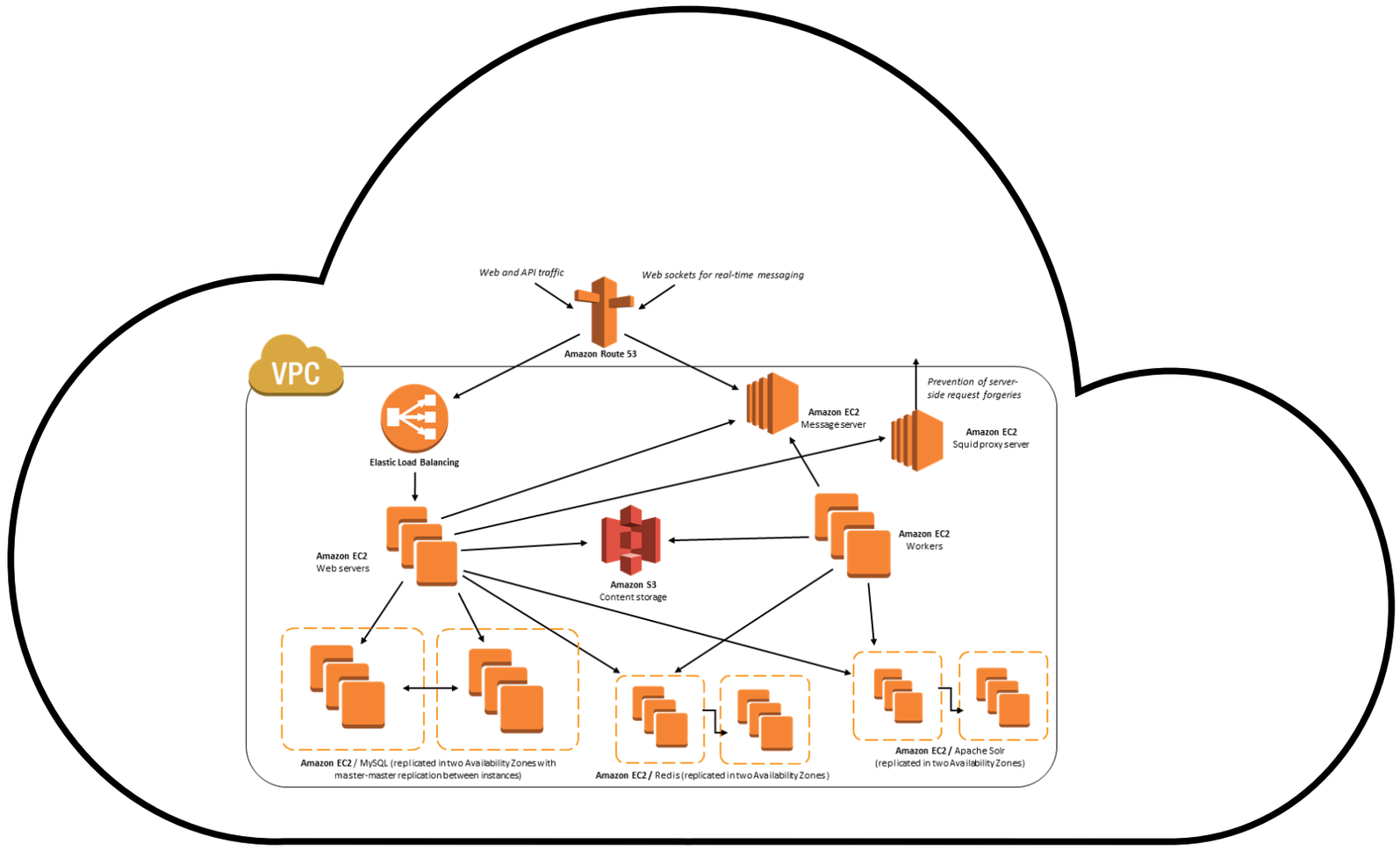
ID Card?

Receipt?

Insurance Contract?

Integration

Where should we put our whole new stack of technology?



Quokky VPC

Brute Force Approach

- How hard can it be to **classify a bunch of black & white pixels** generated by machine, right?

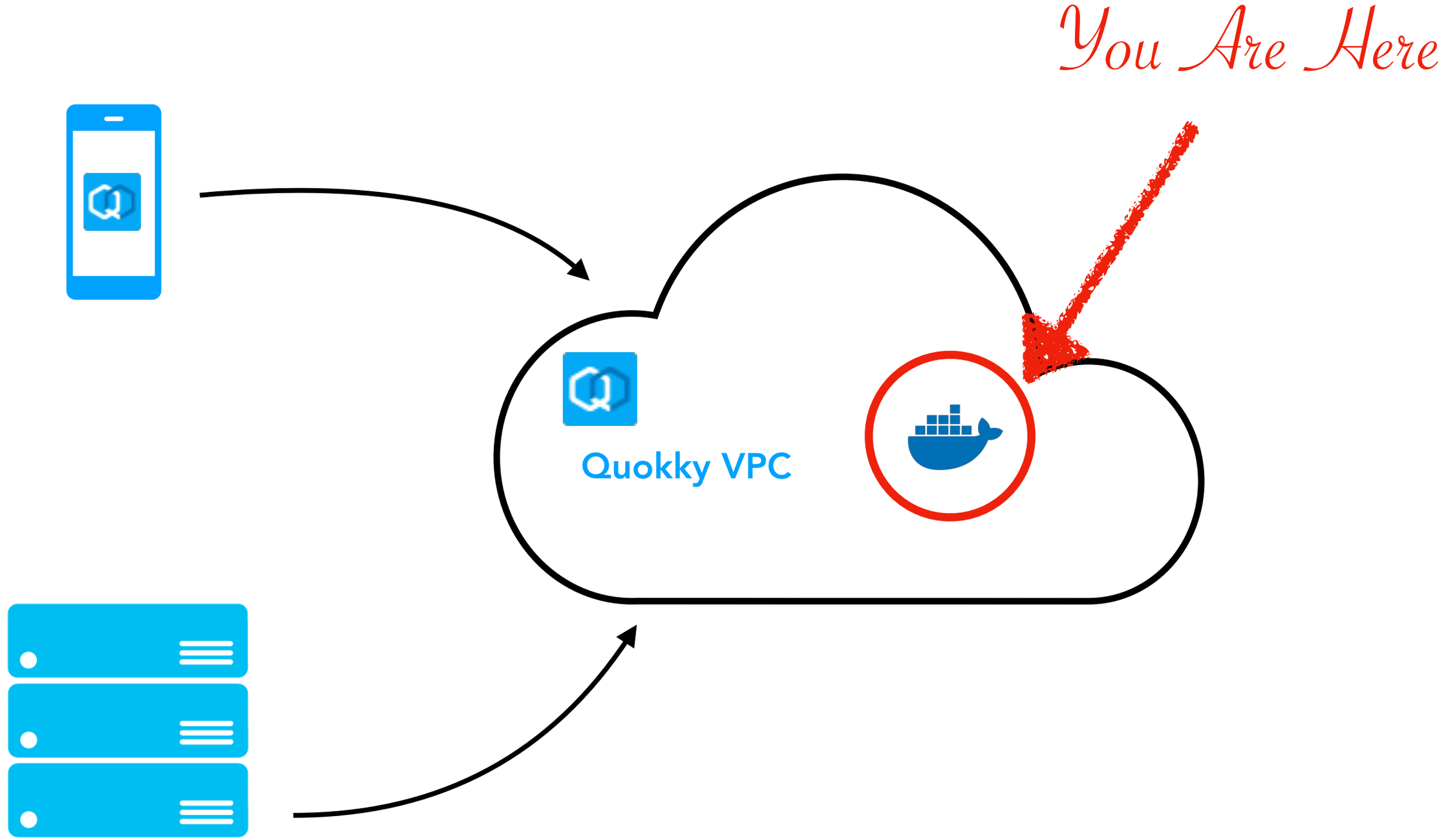
- **Simply** extend our VPC adding another service!

Optical Mark Recognition

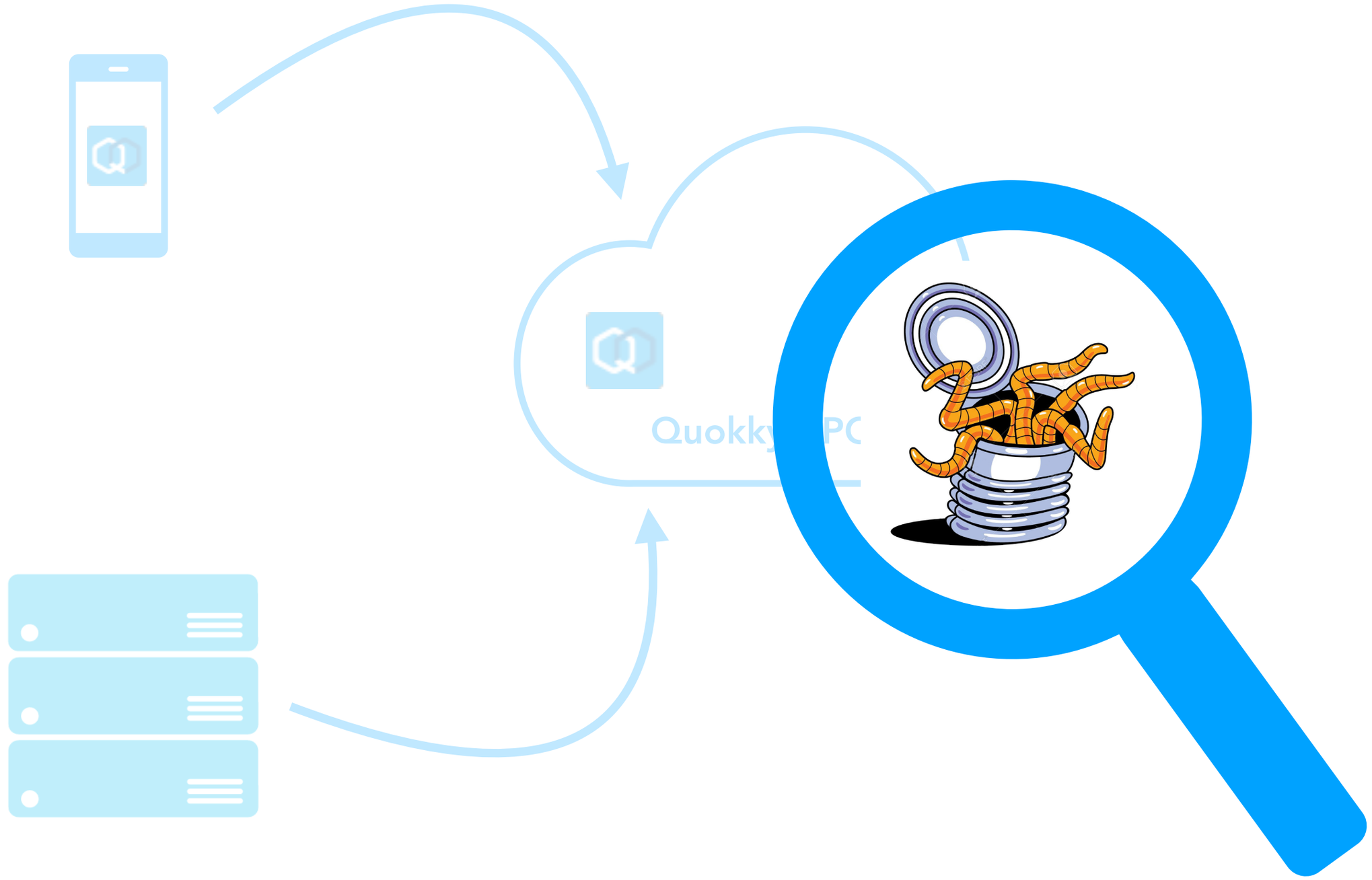
- Chain a **third-party** OCR + image processing tools
- OMR developed in house (*)
  - Bitmap to vector
  - **Manual** Path analysis
- Basically templating (on steroids)

(*) `github.com/apiraino/signature-recognition`

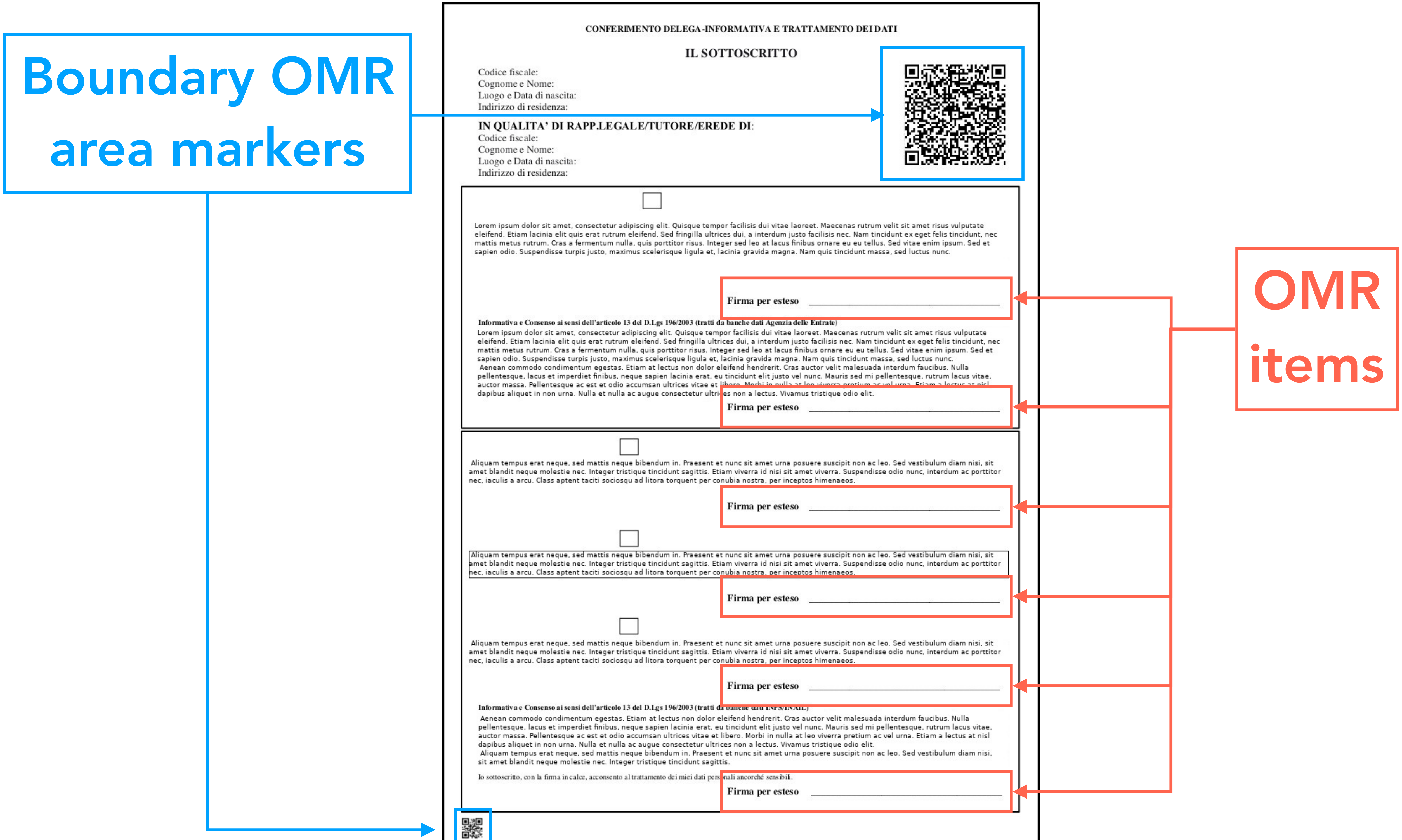
Dockerize it!



Service Architecture overview



OMR


Boundary OMR area markers
OMR items
CONFERIMENTO DELEGA-INFORMATIVA E TRATTAMENTO DEI DATI
IL SOTTOSCRITTO
Codice fiscale:
Cognome e Nome:
Luogo e Data di nascita:
Indirizzo di residenza:
IN QUALITA' DI RAPP.LEGALE/TUTORE/EREDE DI:
Codice fiscale:
Cognome e Nome:
Luogo e Data di nascita:
Indirizzo di residenza:
Firma per esteso
Firma per esteso
Firma per esteso
Firma per esteso
Firma per esteso
Firma per esteso

OMR

PDF preprocessing
+
OCR

QR code recog.
+
resample

OMR workflow

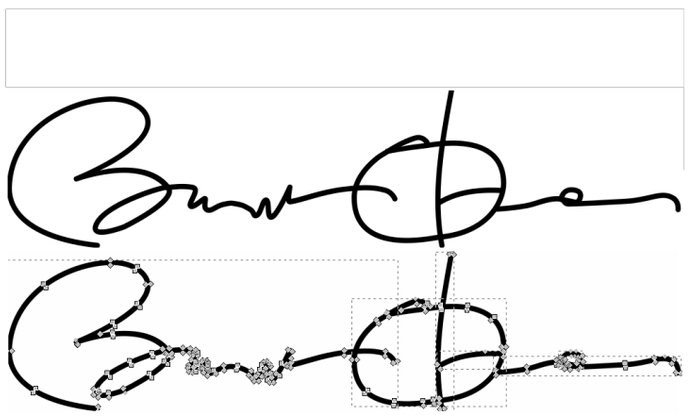
Convert to TIFF B/N Group4

Crop signatures

Convert BMP to SVG

SVG path analysis

row 1: x, y, w, h
row 2: x, y, w, h
row 3: x, y, w, h
...

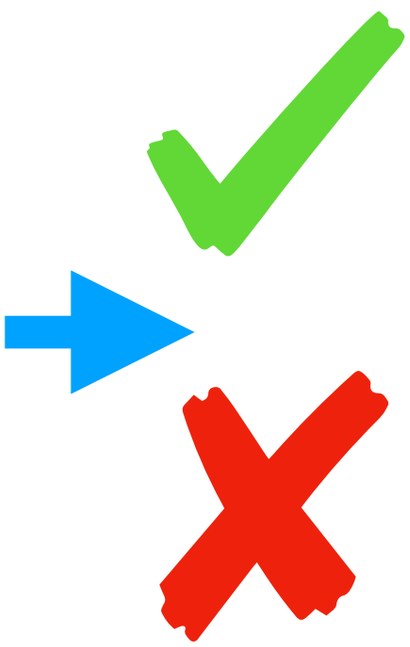


OMR

PDF preprocessing
+
OCR

QR code recog.
+
resample

Do we have a signature?



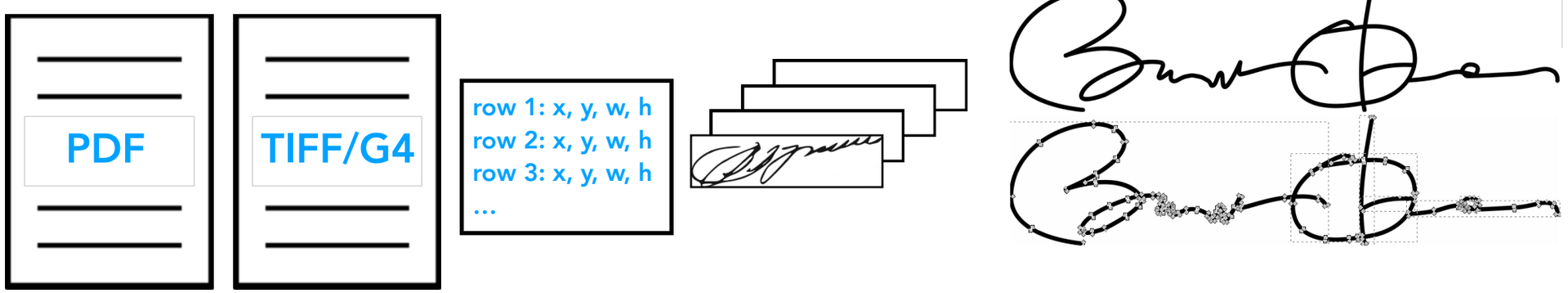
OMR workflow

Convert to TIFF B/N Group4

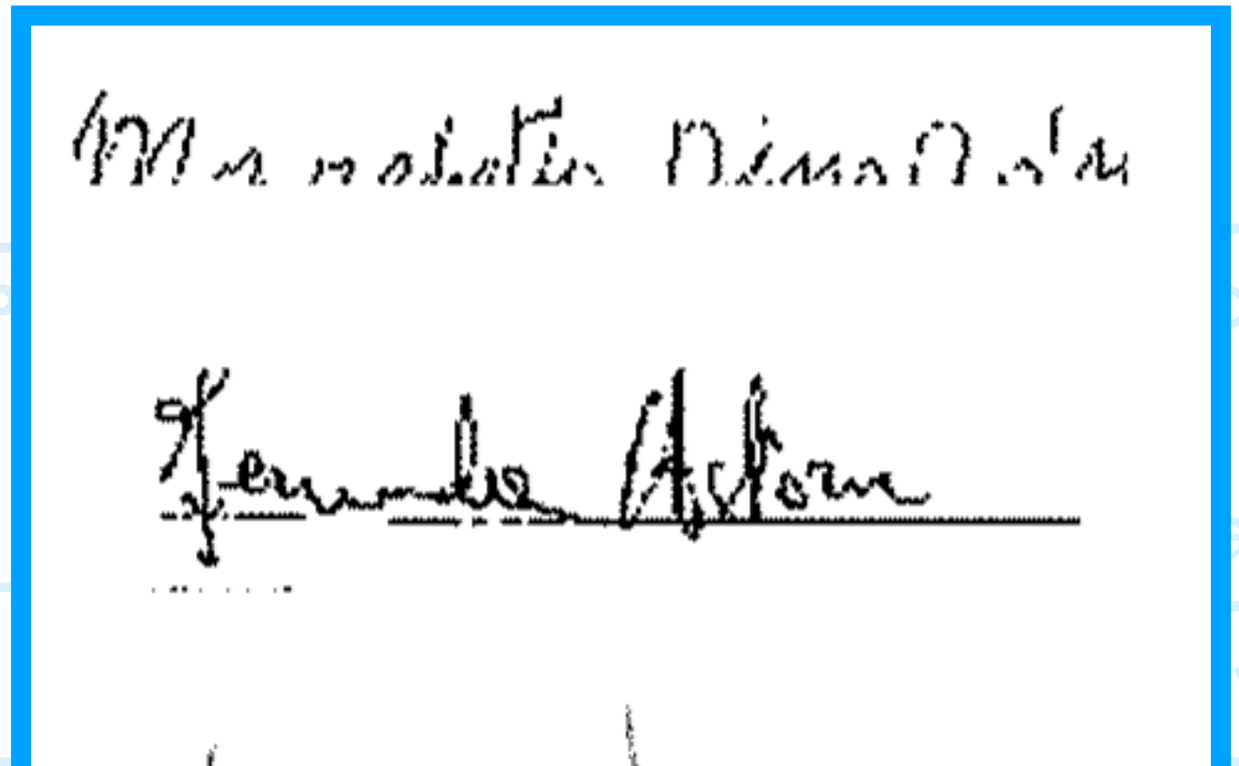
Crop signatures

Convert BMP to SVG

SVG path analysis



OMR



OMR workflow

Convert
to TIFF B/N
Group4

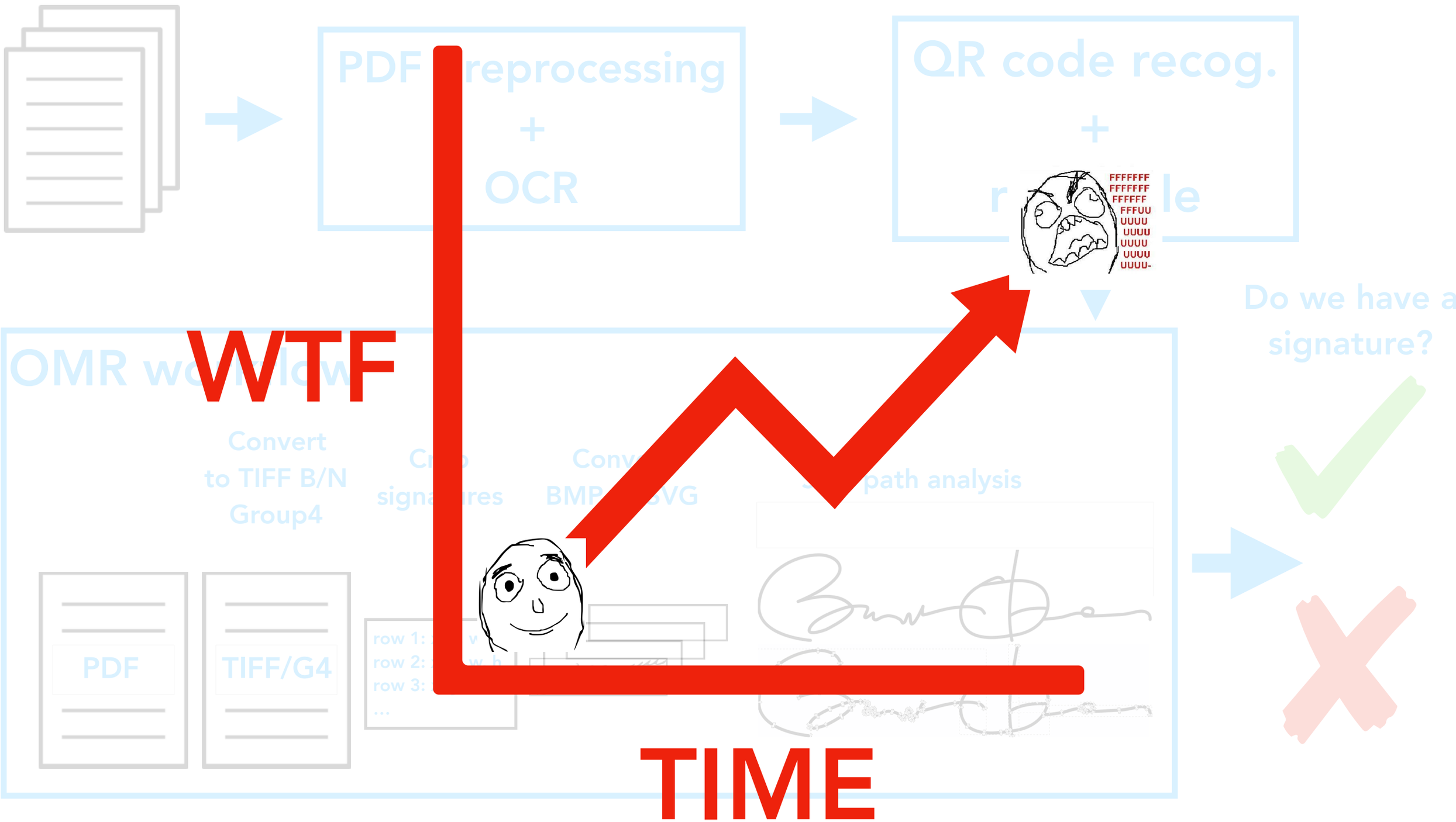
PDF

TIFF/G4

de recog.
+
ample

Do we have a
signature?

WTFs / min.


PDF reprocessing
+
OCR
QR code recog.
+
r le
FFFFFFF
FFFFFFF
FFFFFF
FFFUU
UUUU
UUUU
UUUU
UUUU
UUUU-
Do we have a signature?
OMR w low
Convert to TIFF B/N Group4
signatures
BMP SVG
path analysis
PDF
TIFF/G4
row 1:
row 2:
row 3:
...
WTF
TIME

Fake Confidence

```
if _signed == 'Y':
    if signature_rating == RATING_GOOD:
        _confidence += 0.3
    elif signature_rating == RATING_WARN:
        _confidence -= 0.1
    elif signature_rating == RATING_BAD:
        _confidence -= 0.2
    elif signature_rating == RATING_UNKNOWN:
        _confidence -= 0.3
else:
    if signature_rating == RATING_GOOD:
        _confidence += 0.2
    elif signature_rating == RATING_WARN:
        _confidence -= 0.1
    elif signature_rating == RATING_BAD:
        _confidence += 0.4
    elif signature_rating == RATING_UNKNOWN:
        _confidence += 0.4
return _confidence
```

OMR Results

+ deployed a *beta* in production
+ worked pretty well for machine-generated docs
+ could even handle some manually-generated content

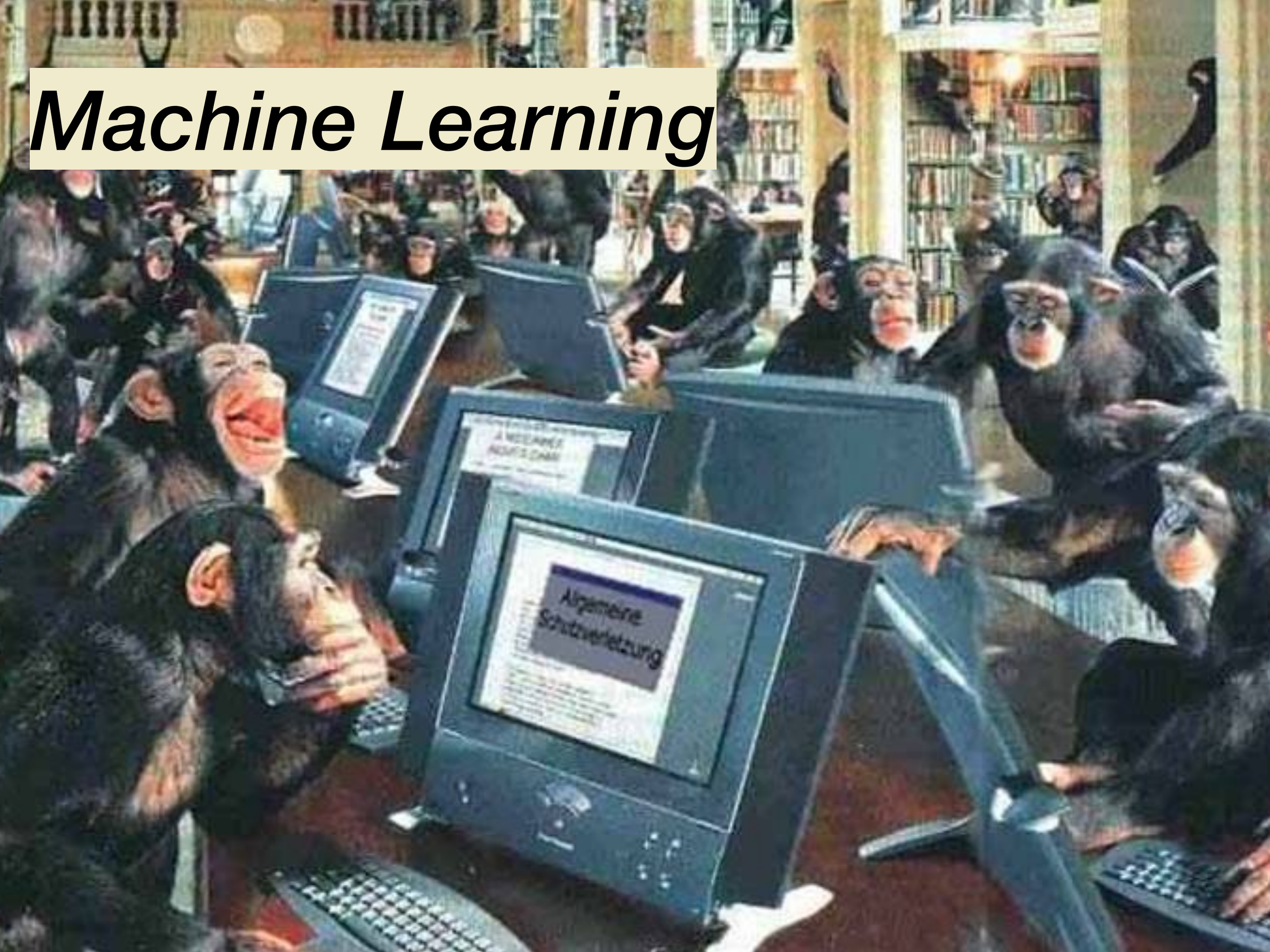
- no real training done (thresholds finetune) b/c …
- … good luck explaining what a “test dataset” is
- weak decision tree
- not flexible
- a mess of external FOSS and internal tools
- licensing fees $$$, SDK bugs, bad dev support (*)

* Yes, **paid** support

What we learned

- when all you have is a hammer …
- the dangers of “quick” prototyping
- no single approach will work for everything

Machine Learning



Roadmap

ML Glossary
To get everyone on same page

Computer Vision and CNN
How a computer sees our world?
Brief intro to Convolutional Neural Network & TensorFlow

Custom Image Classifier (CIC)
What is Inception Model and how to retrain it using TensorFlow?

CIC's use case
Signatures & Document classifier

Textual Classifier
Categorise visually identical documents using Amazon ML

ML Glossary

Dataset (or datasource): Your input data (texts, images) and desired output category

Supervised learning: If you know your dataset and you can label your data

Training Phase: Iterating over dataset to find **patterns**

ML Model: Stats related to probability distributions

Evaluation: Assert how good the trained model is

ML Solution Workflow

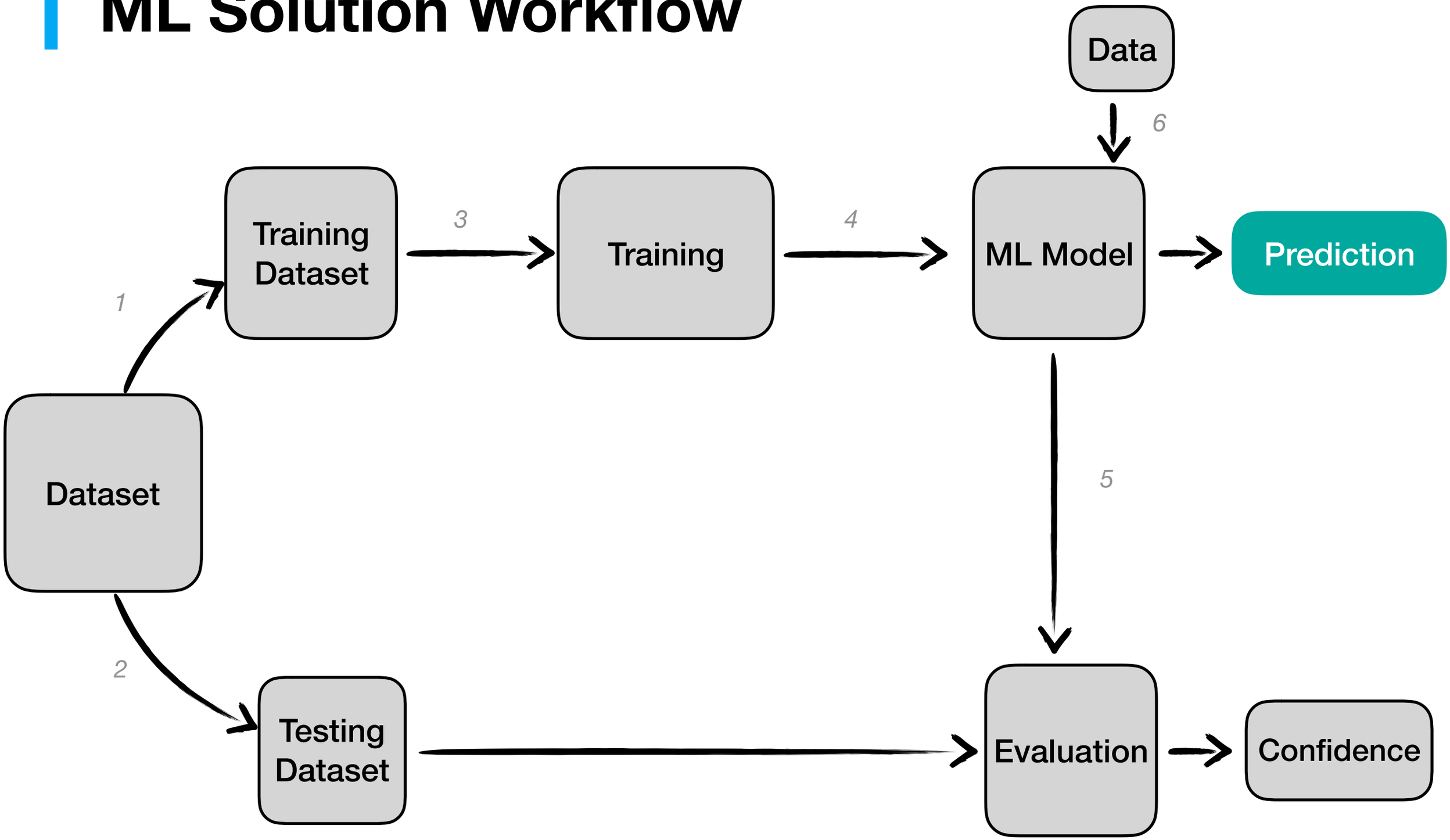
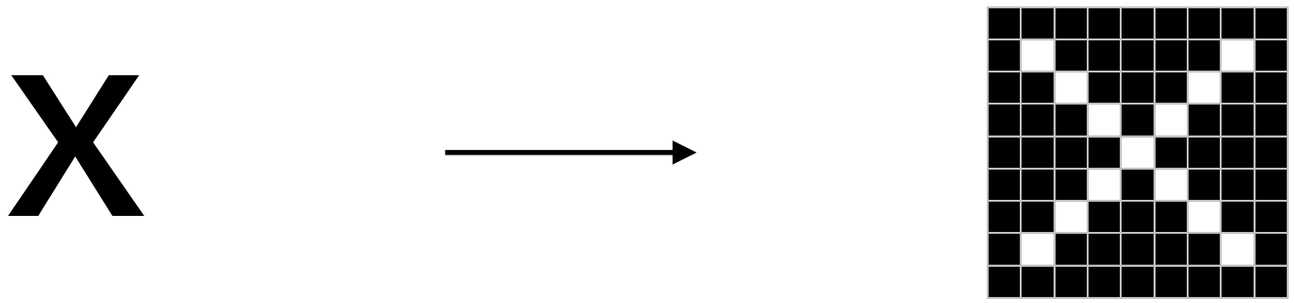


Image is Array of numbers

Each image can be thought as 2-dimensional array of numbers

We can consider black as 0 and White as 1



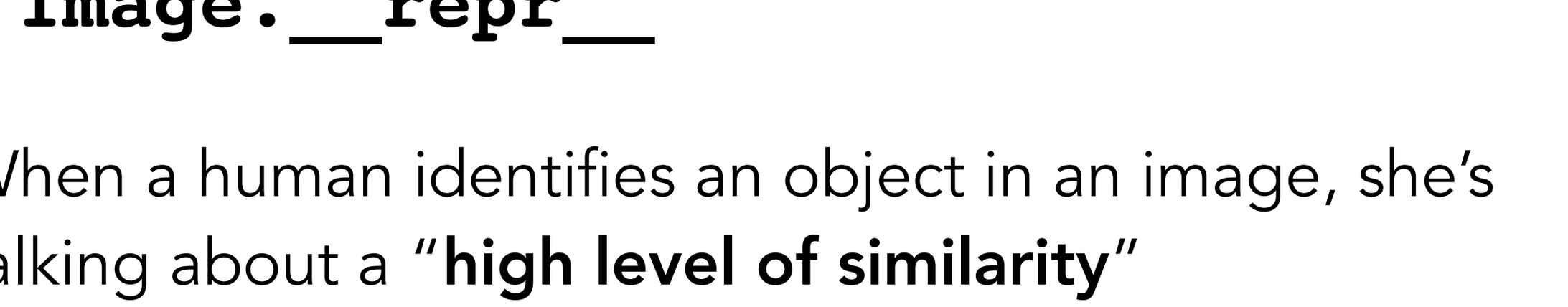
`Image.__repr__`

When a human identifies an object in an image, she's talking about a "**high level of similarity**"

`Image.__repr__`

When a human identifies an object in an image, she's talking about a "**high level of similarity**"

Comparing pixels one by one will fail, because **same** object can be translated, scaled, rotated, shifted:

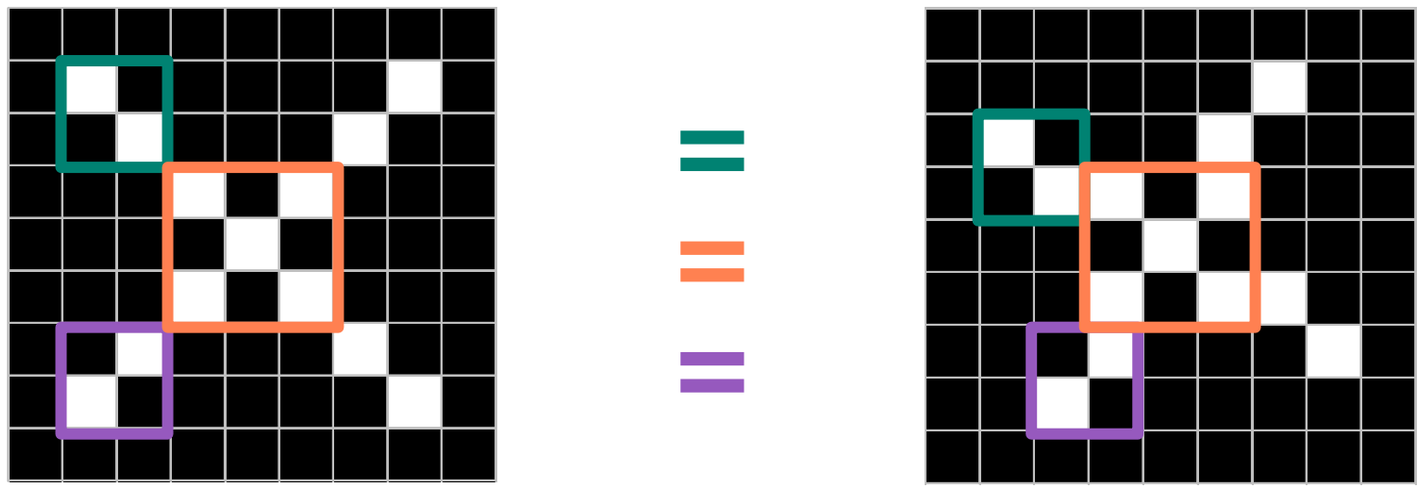


Convolutional Network

Set of (simple math) rules to find **patterns** in an array.

Replicate "**experience**" for the computer.

Single pixel will loose its importance in an image, and instead the **pieces of the image** will define what is in the picture.



Tools

Filtering

Pooling

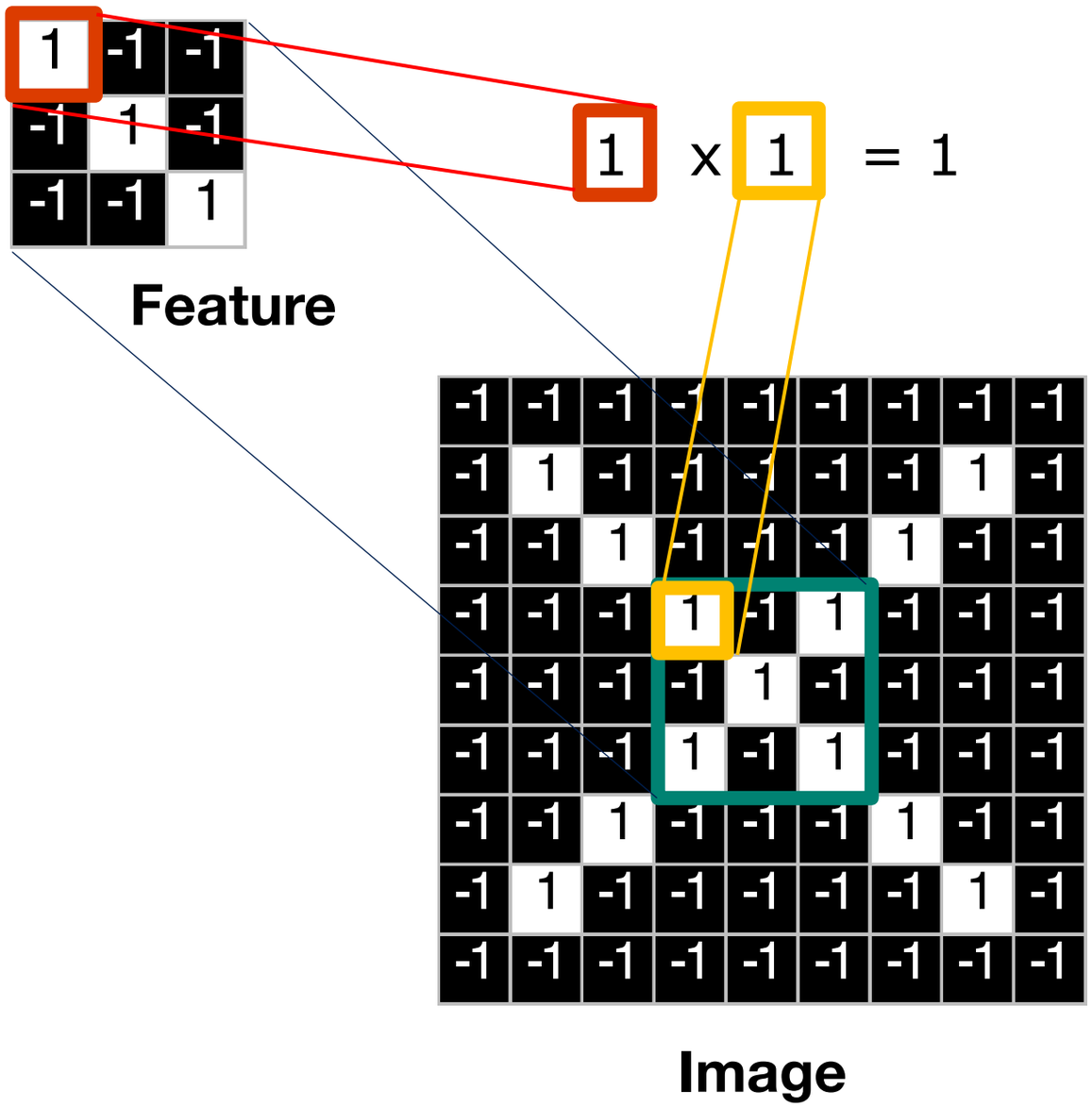
Normalising

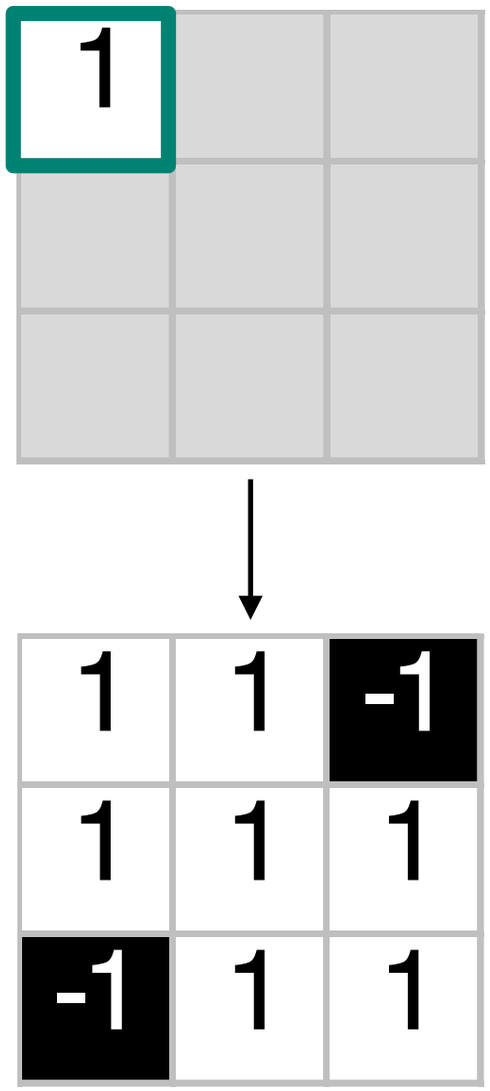
Fully Connected Network

Filtering

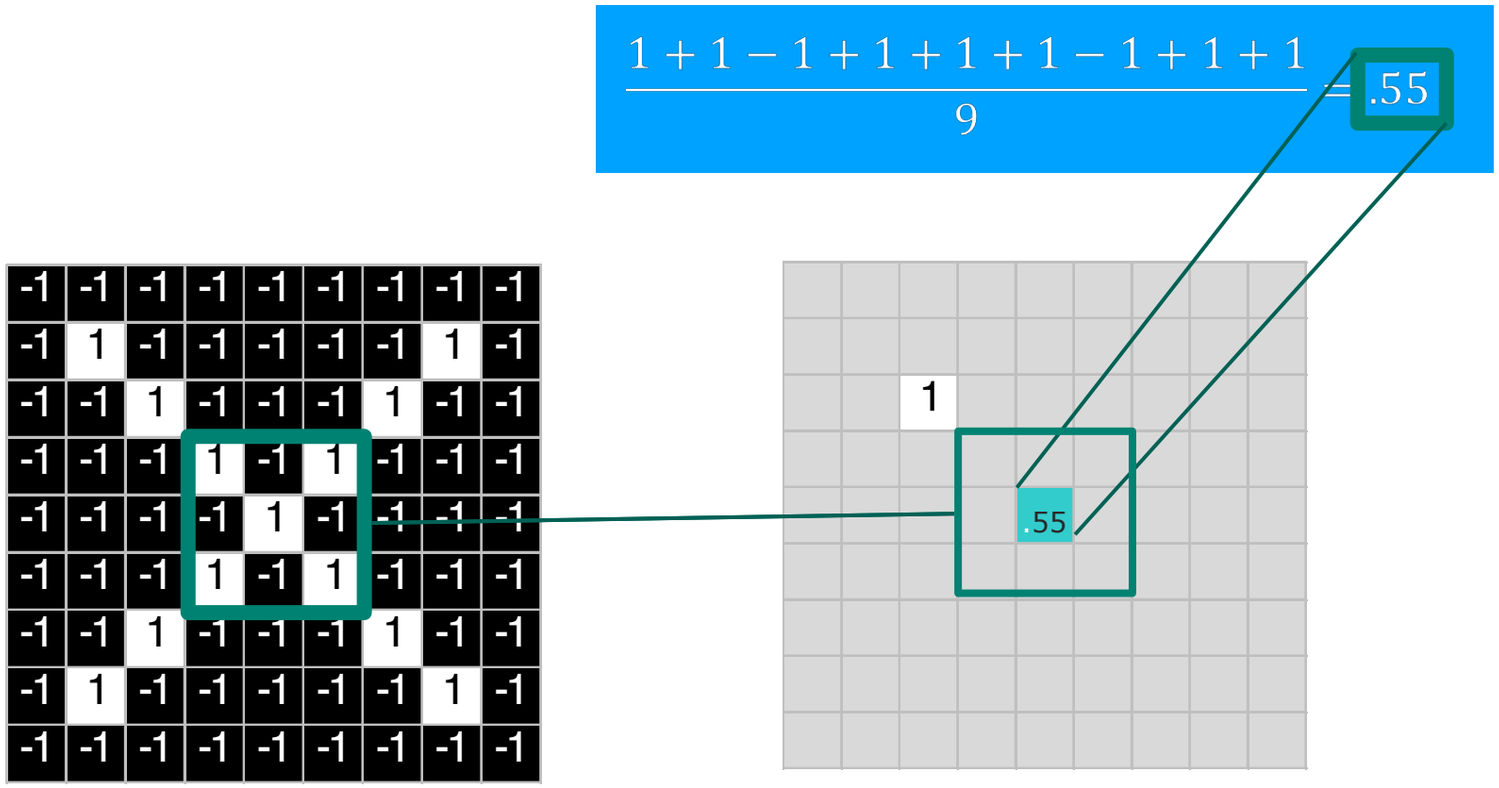
Replace a set of pixels (**the feature**) with a single float value

Filtering


1 -1 -1
-1 1 -1
-1 -1 1
Feature
1 x 1 = 1
-1 -1 -1 -1 -1 -1 -1 -1 -1
-1 1 -1 -1 -1 -1 -1 1 -1
-1 -1 1 -1 -1 -1 1 -1 -1
-1 -1 -1 1 -1 1 -1 -1 -1
-1 -1 -1 -1 1 -1 -1 -1 -1
-1 -1 -1 1 -1 1 -1 -1 -1
-1 -1 1 -1 -1 -1 1 -1 -1
-1 1 -1 -1 -1 -1 -1 1 -1
-1 -1 -1 -1 -1 -1 -1 -1 -1
Image

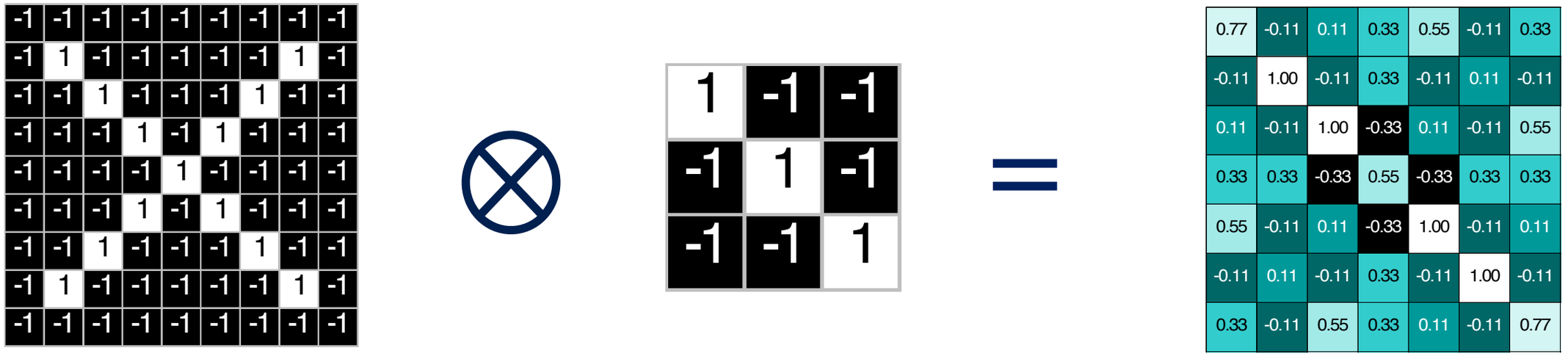

1
1 1 -1
1 1 1
-1 1 1

Filtering


$\frac{1+1-1+1+1+1-1+1+1}{9} = .55$
-1 -1 -1 -1 -1 -1 -1 -1 -1
-1 1 -1 -1 -1 -1 -1 1 -1
-1 -1 1 -1 -1 -1 1 -1 -1
-1 -1 -1 1 -1 1 -1 -1 -1
-1 -1 -1 -1 1 -1 -1 -1 -1
-1 -1 -1 1 -1 1 -1 -1 -1
-1 -1 1 -1 -1 -1 1 -1 -1
-1 1 -1 -1 -1 -1 -1 1 -1
-1 -1 -1 -1 -1 -1 -1 -1 -1
1
.55

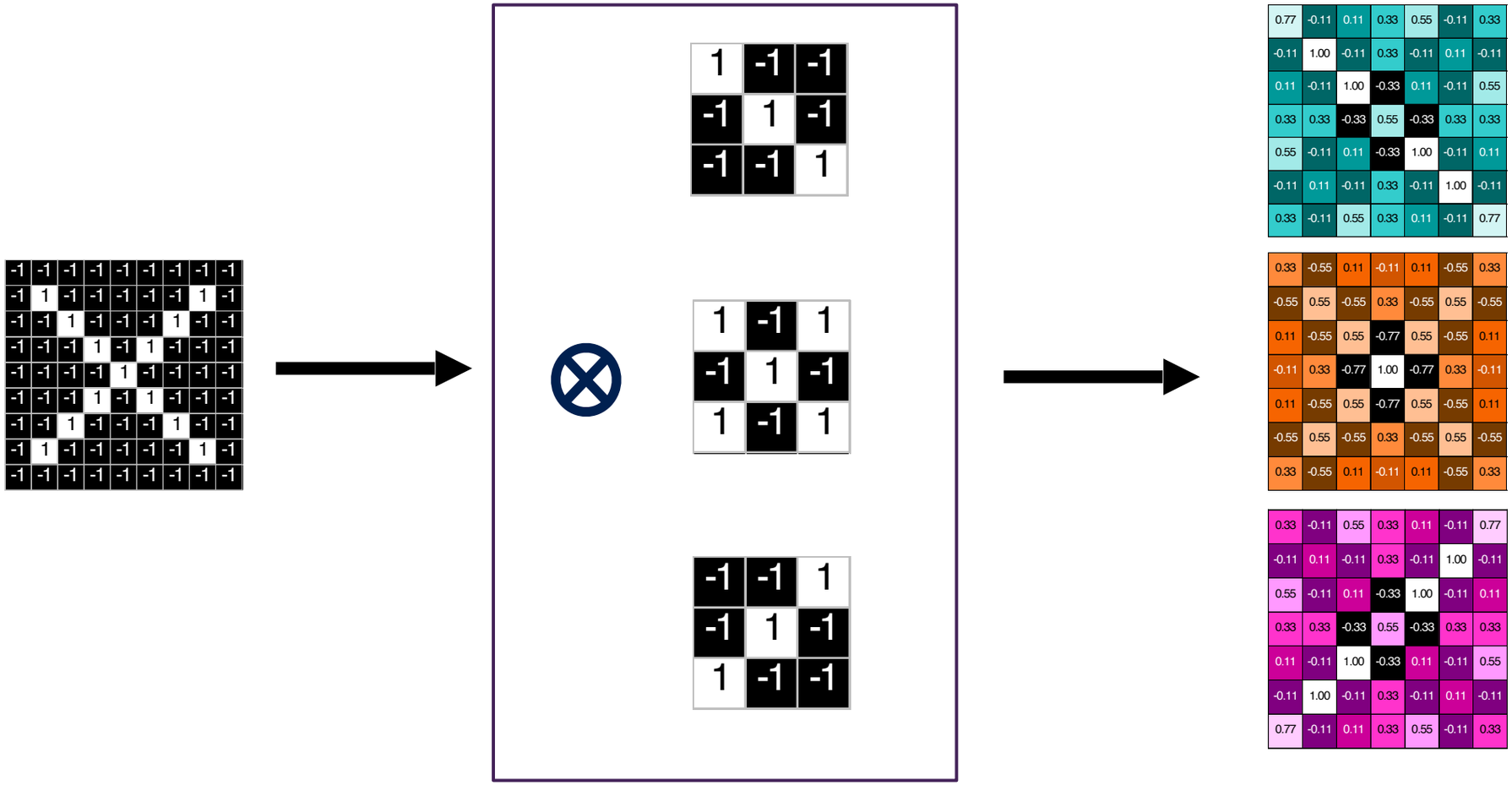
Filtering

By moving around a filter to **every possible position** we would have **smaller image** with **map** of where the feature occurs.



Convolutional Layer

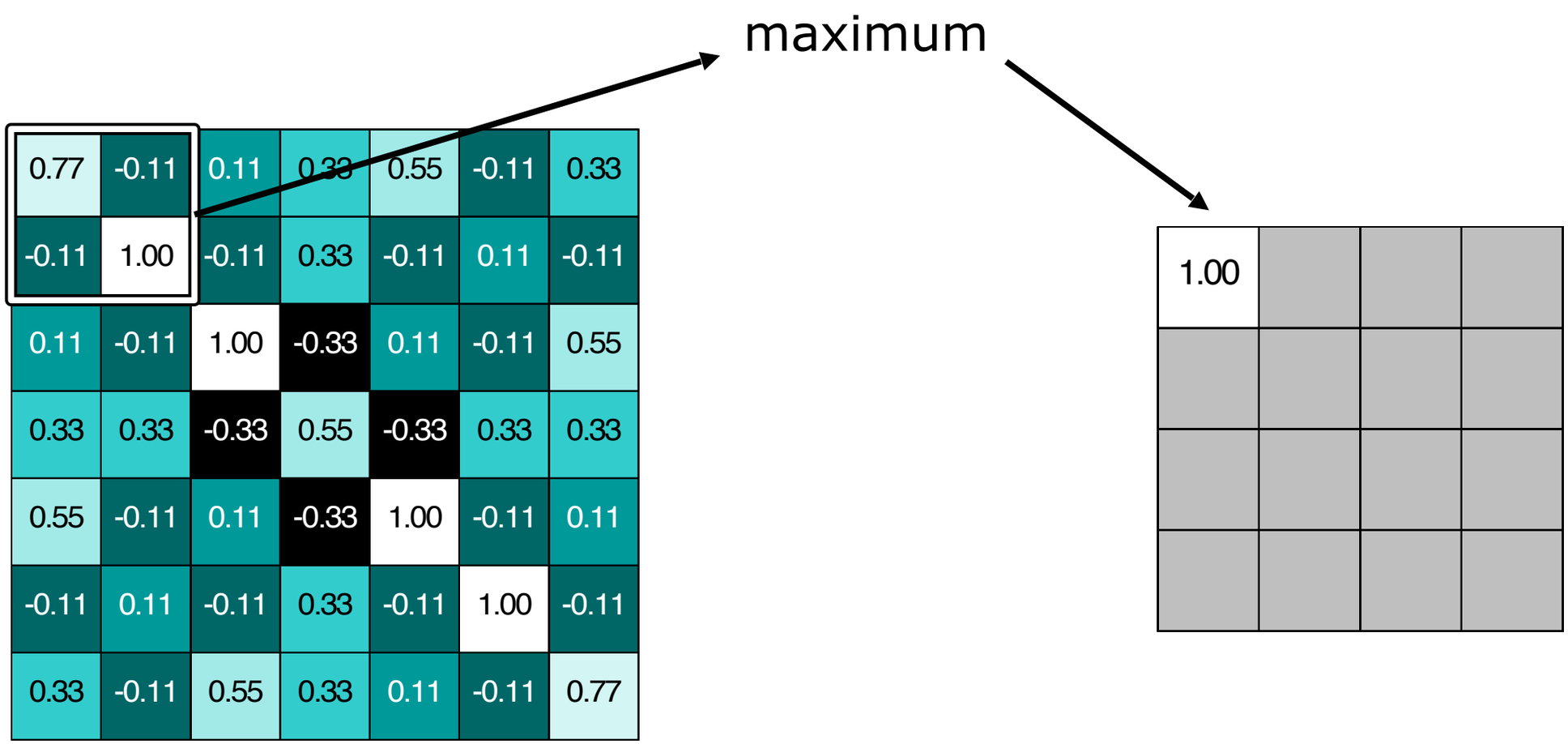
One image becomes a stack of filtered images



Pooling

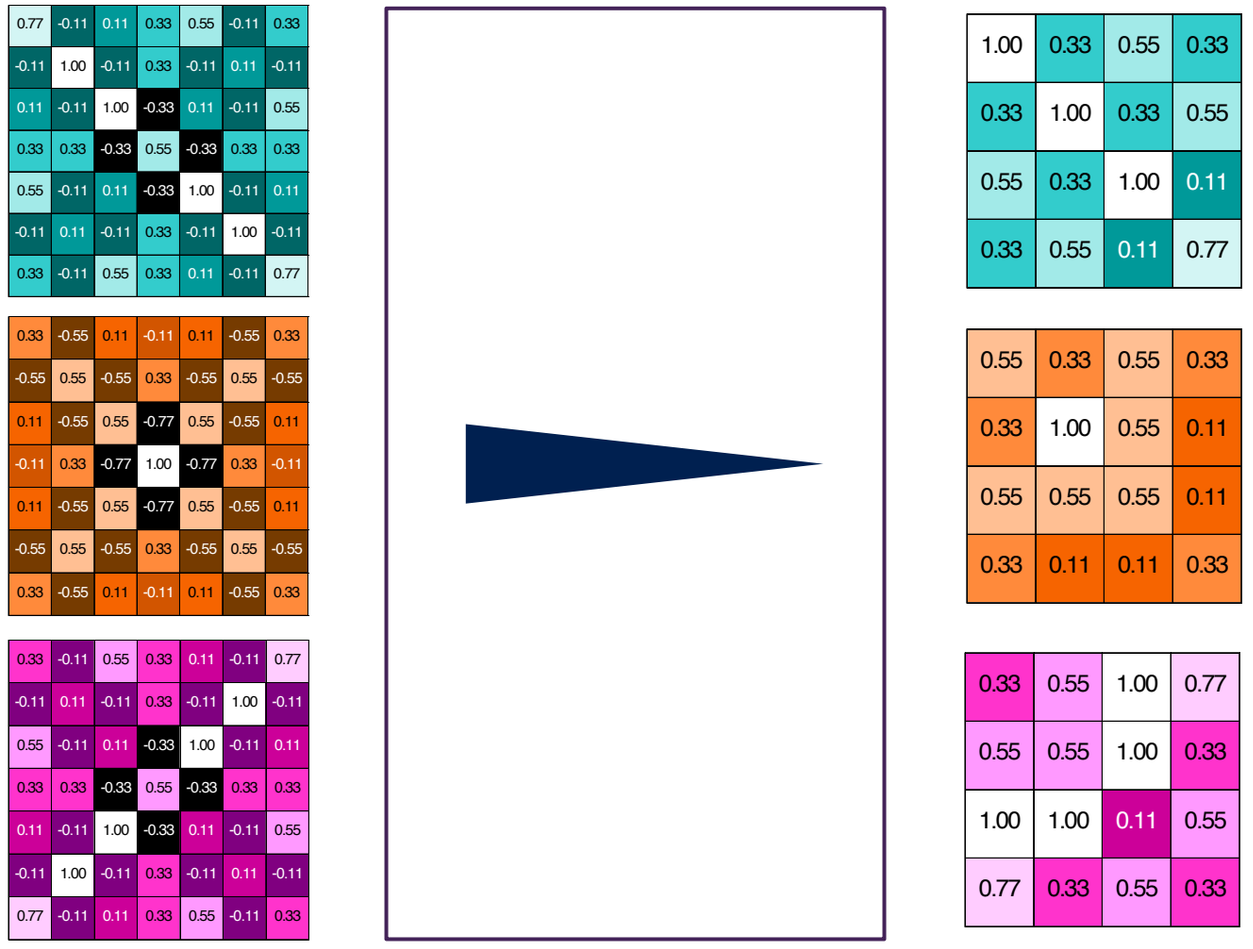
Window **shrinking** by aggregating most prominent features

(max) Pooling


maximum
0.77 -0.11 0.11 0.33 0.55 -0.11 0.33
-0.11 1.00 -0.11 0.33 -0.11 0.11 -0.11
0.11 -0.11 1.00 -0.33 0.11 -0.11 0.55
0.33 0.33 -0.33 0.55 -0.33 0.33 0.33
0.55 -0.11 0.11 -0.33 1.00 -0.11 0.11
-0.11 0.11 -0.11 0.33 -0.11 1.00 -0.11
0.33 -0.11 0.55 0.33 0.11 -0.11 0.77
1.00

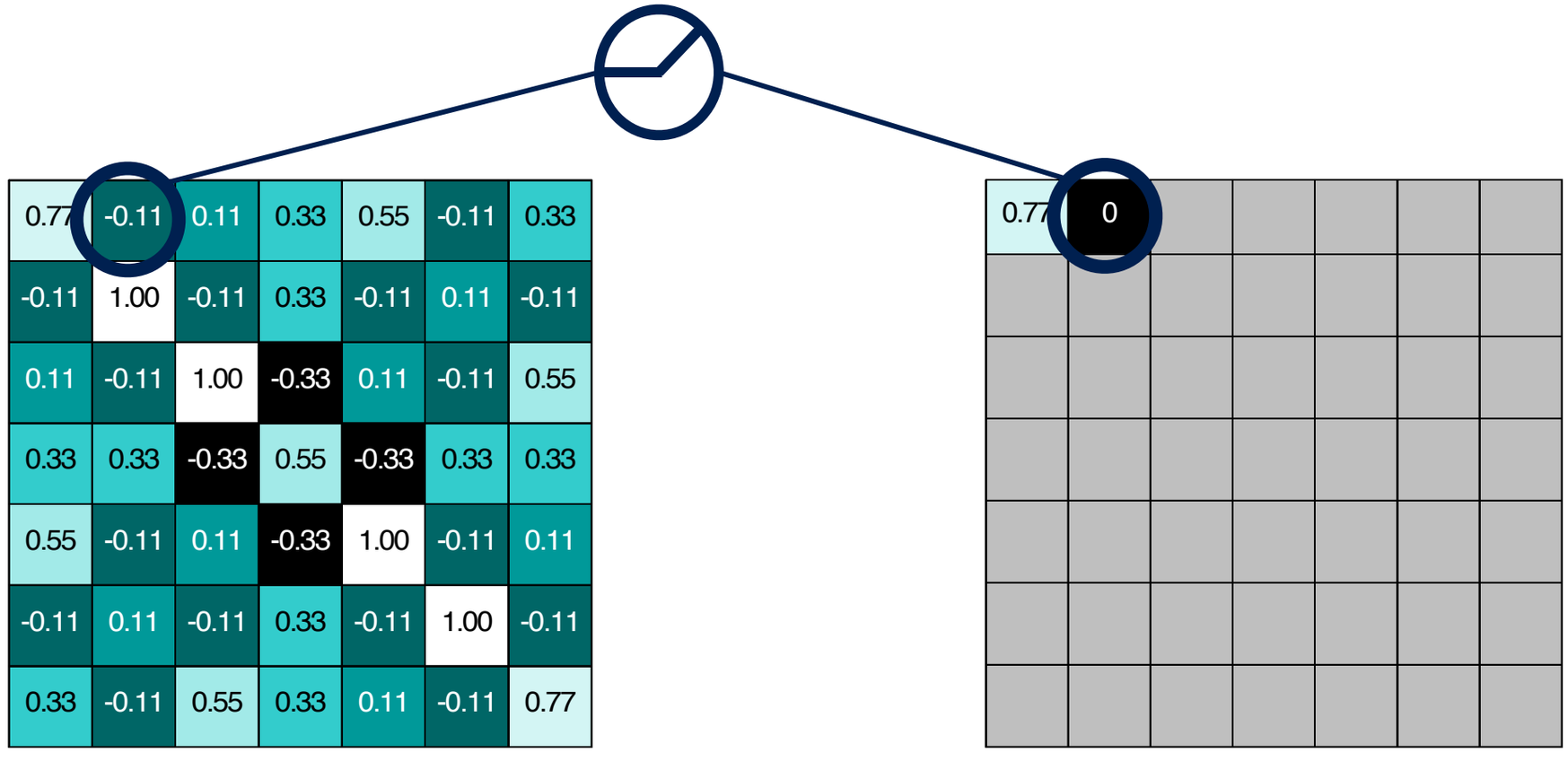
Pooling

A stack of images becomes a stack of smaller images.



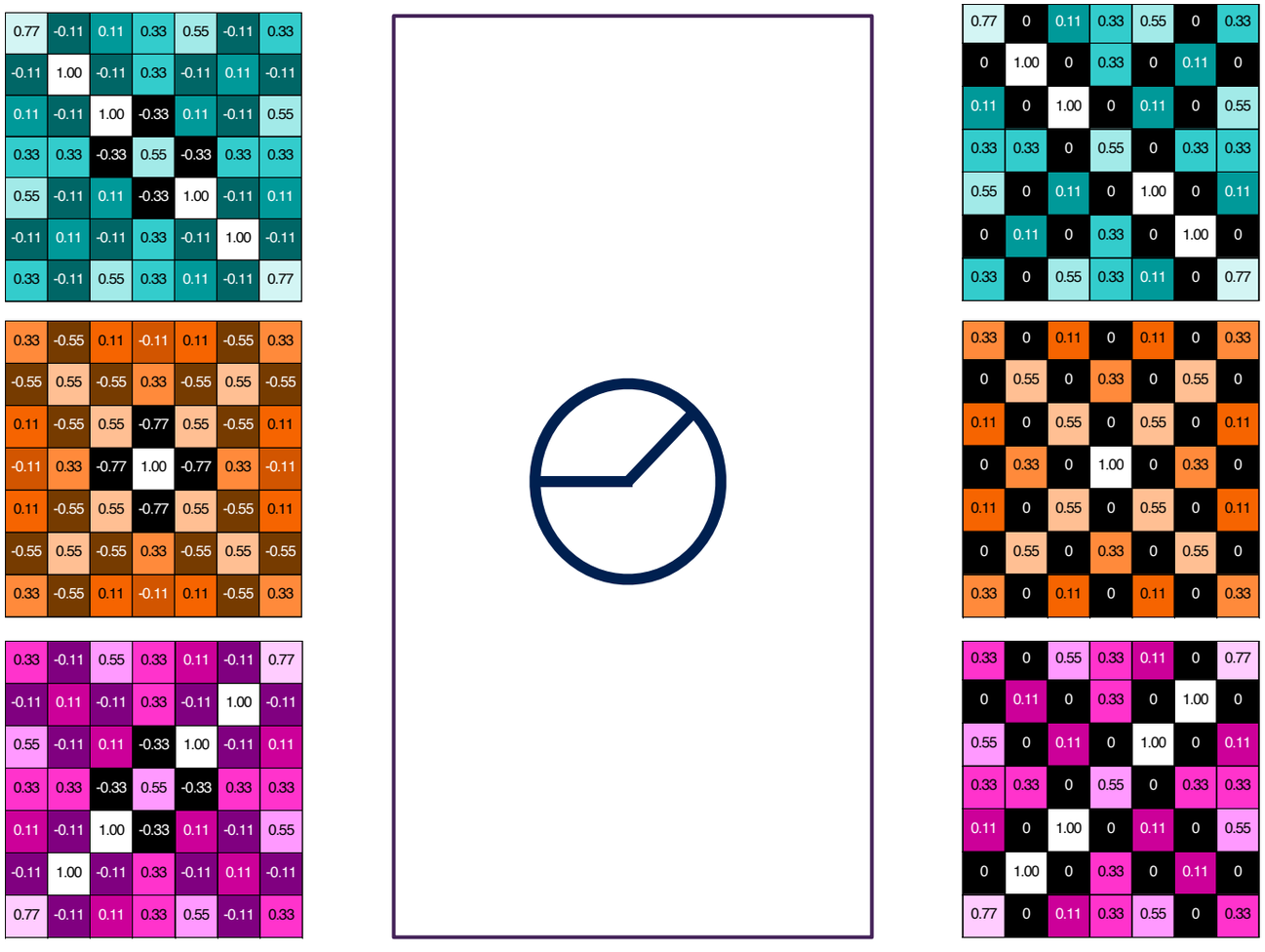
Normalisation

Change everything negative to zero.
aka Rectified Linear Units (ReLUs)

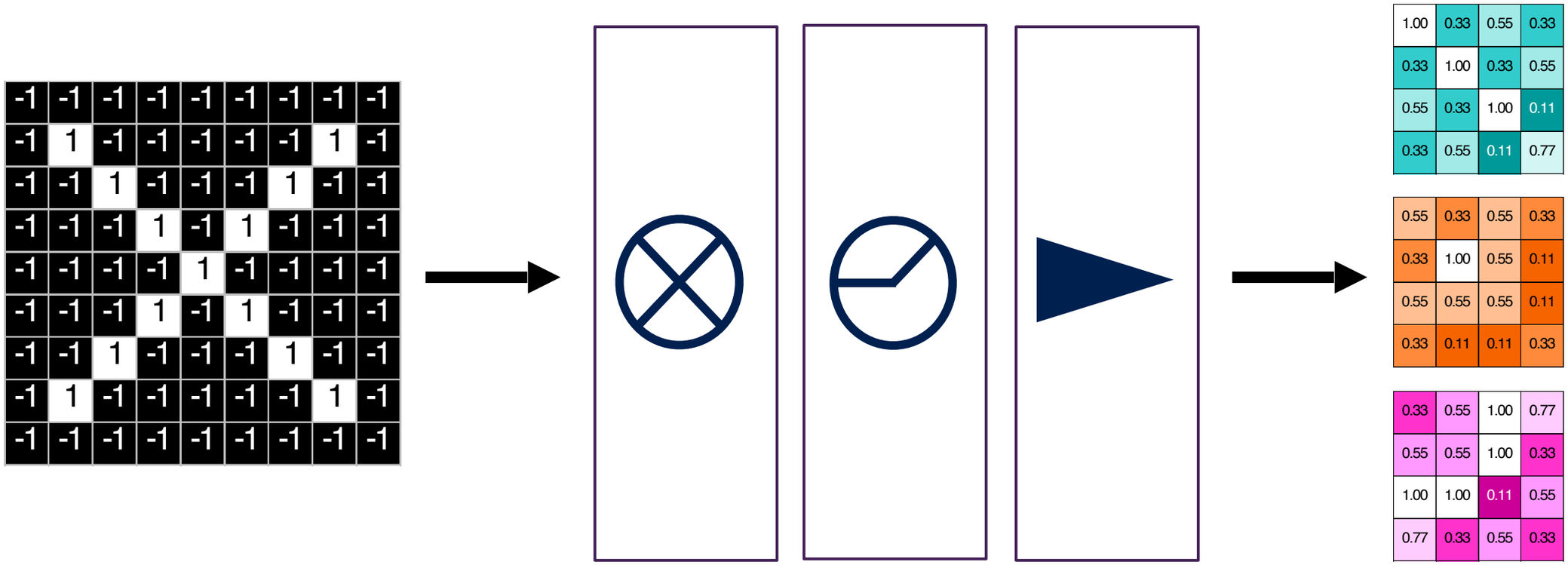


ReLU Layer

A stack of images becomes a stack of images with no negative values.

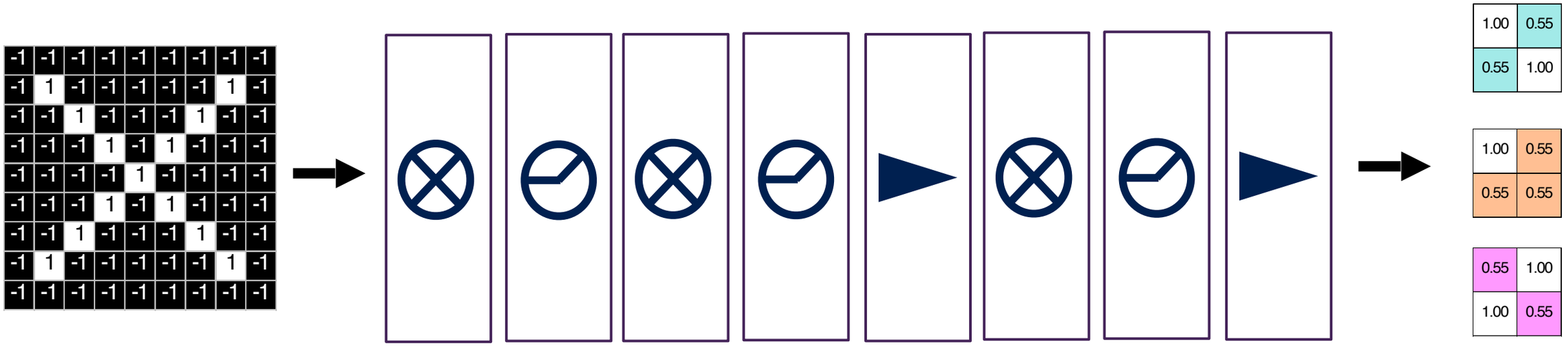


Put everything together



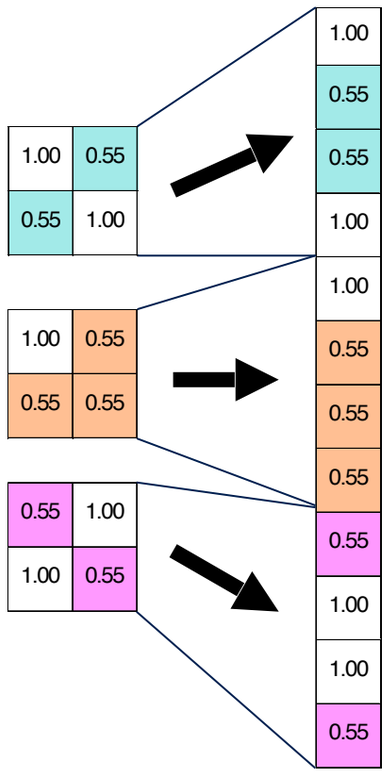
Deep stacking

Layers can be repeated several (or many) times.



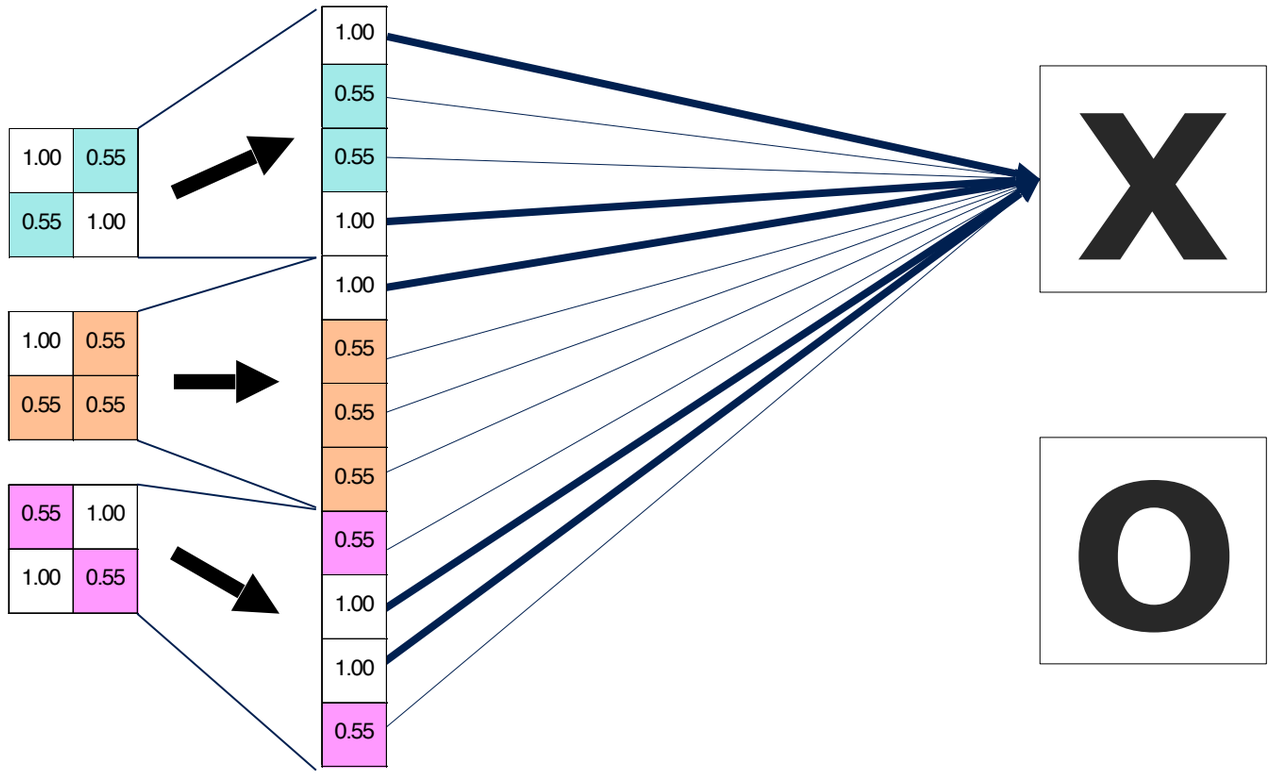
Fully connected layer

Every value gets a vote



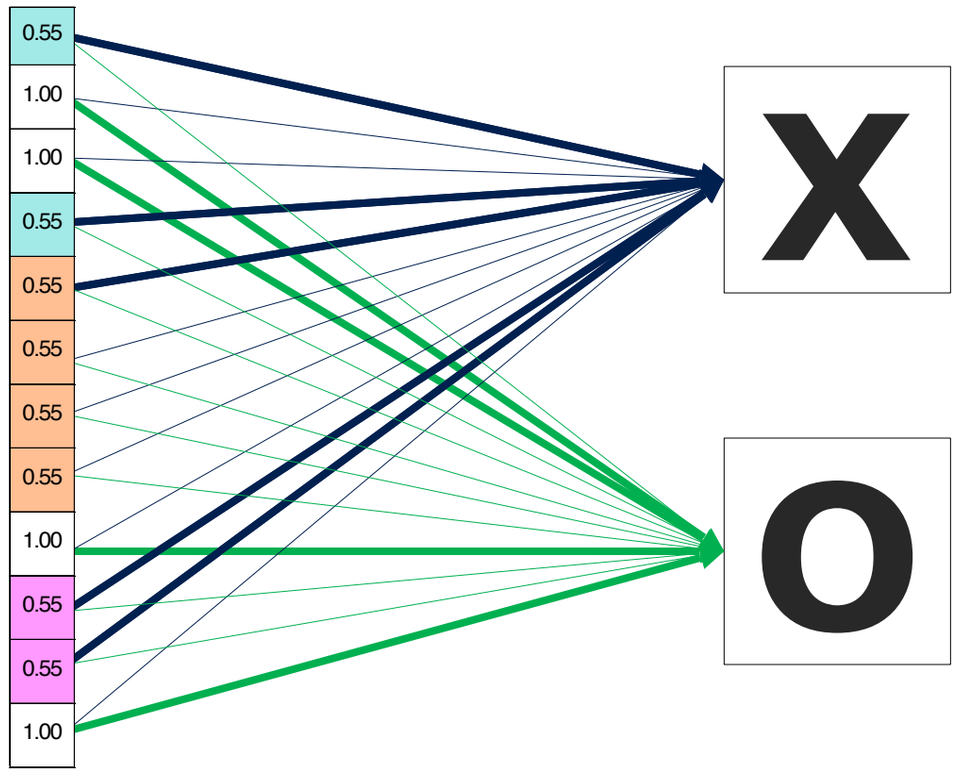
Fully connected layer

Patterns for **X** are repeated on certain points

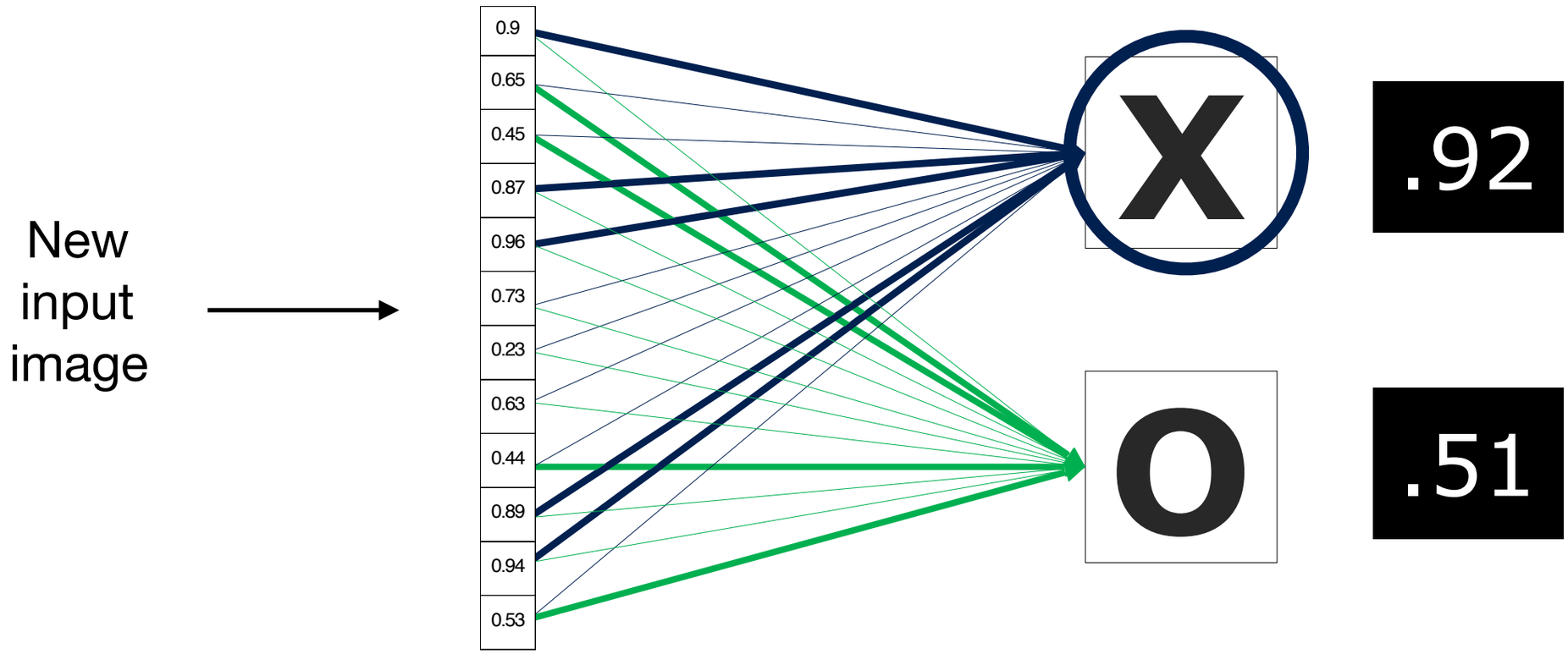


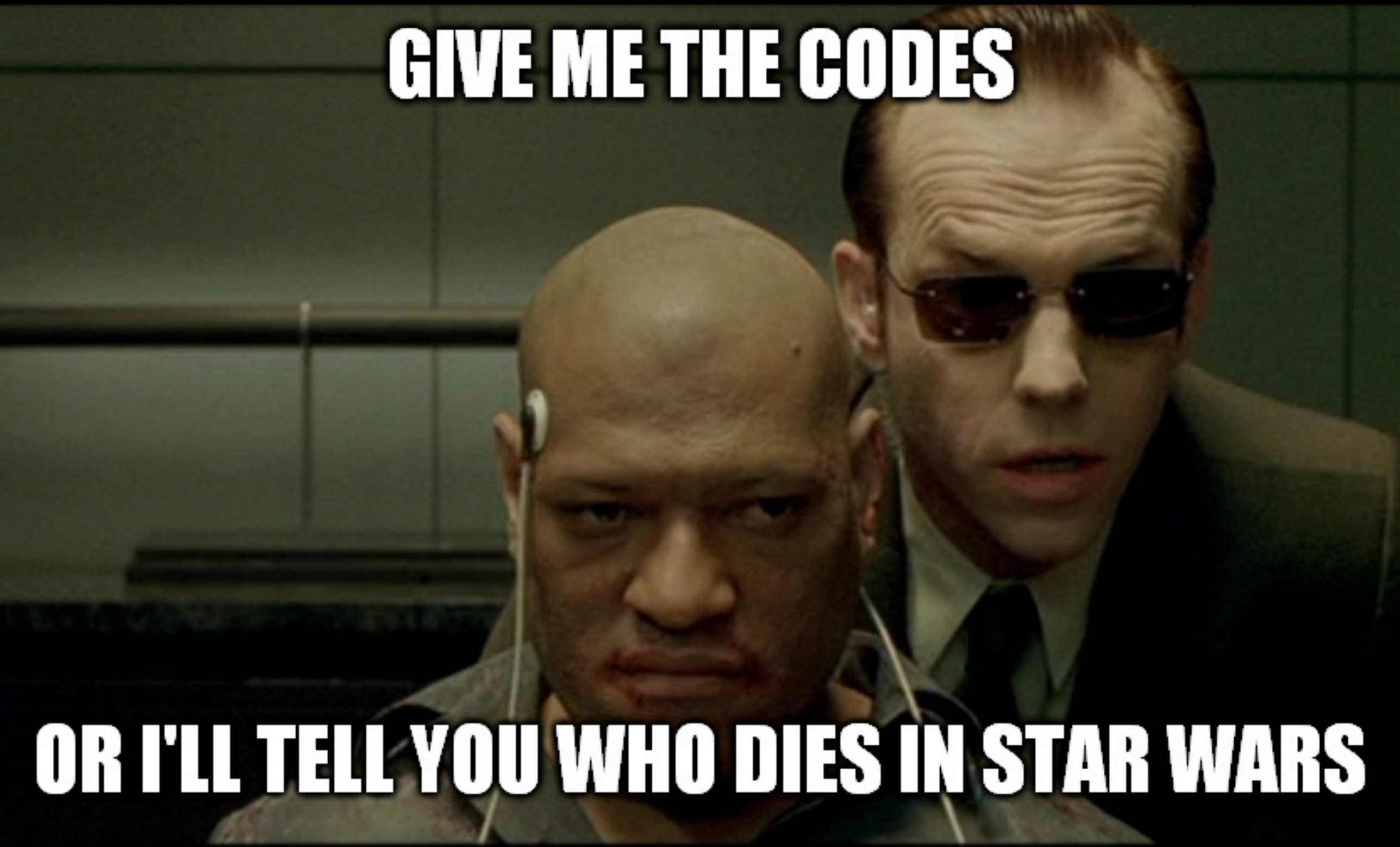
Fully connected layer

Features with higher values have more weight



Predict new image




GIVE ME THE CODES
OR I'LL TELL YOU WHO DIES IN STAR WARS

TensorFlow

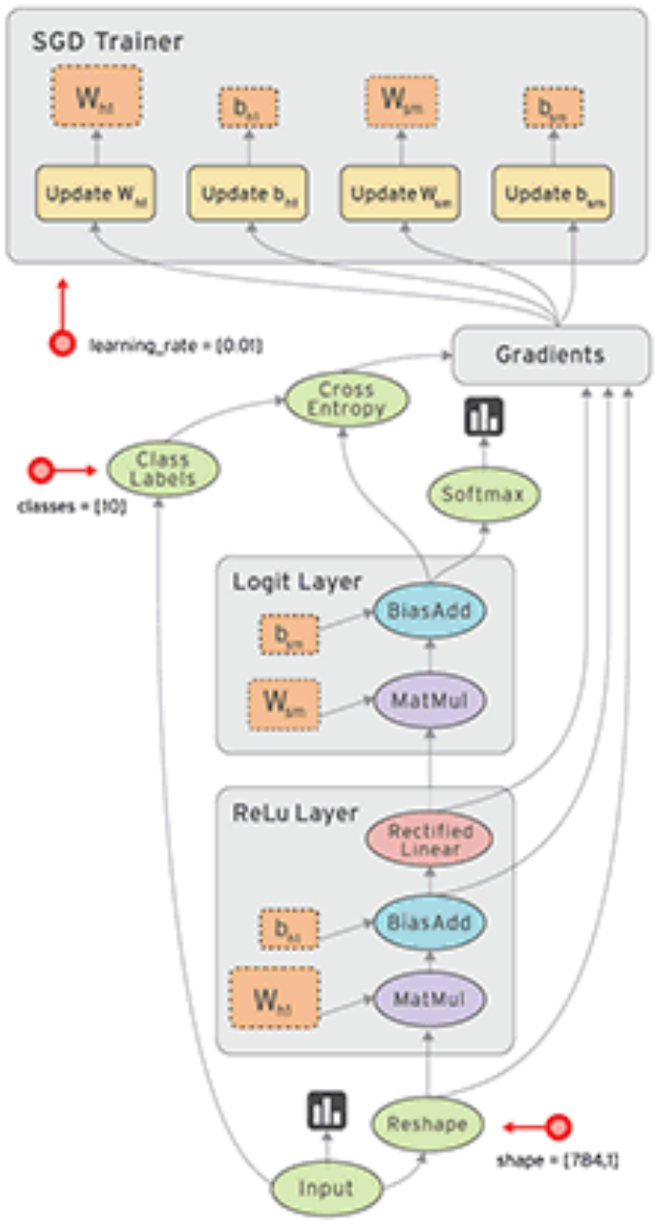
An open source machine learning framework for **everyone** (Optimised for GPU & CPU)

TF's Big Idea

Express a numeric computation as a **Graph**

Why Graphs?

Build **complex** models
using **simple** operations
with **parallel** computation
on **distributed** systems


SGD Trainer
W_{hl}
b_{hl}
W_{sm}
b_{sm}
Update W_{hl}
Update b_{hl}
Update W_{sm}
Update b_{sm}
learning_rate = [0.01]
Gradients
Cross Entropy
Class Labels
classes = [10]
Softmax
Logit Layer
BiasAdd
b_{sm}
W_{sm}
MatMul
ReLu Layer
Rectified Linear
BiasAdd
b_{hl}
W_{hl}
MatMul
Reshape
shape = [784,1]
Input

```
pip install tensorflow
```

tf.Tensor: Basic data structure in TensorFlow.

tf.Operation: Performs computation on tensors.

tf.Graph: Computational graph constructed by tf.Operation (node) and tf.Tensor (edge)

tf.Session: Execute a tf.Graph operations in C++ runtime

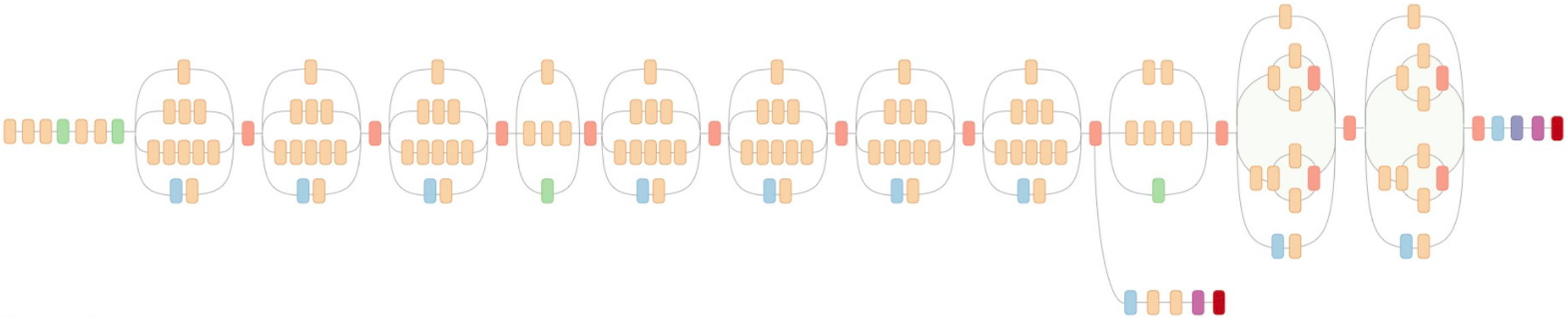
New—**Eager execution**: Evaluates operations immediately

ImageNet Challenge

Over **14 million** of **hand-annotated** images database designed for use in visual object recognition


WE NEED TO GO
DEEPER

Inception V3 Model



Open sourced on 2016

https://research.googleblog.com/2016/03/train-your-own-image-classifier-with.html

Rethinking the Inception Architecture for Computer Vision

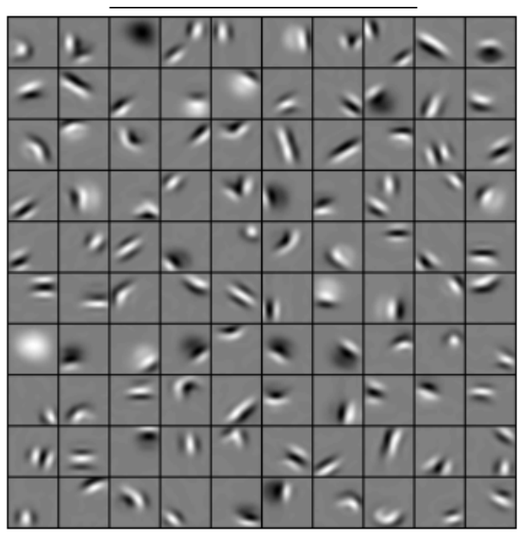
https://arxiv.org/abs/1512.00567

Transfer Learning

Storing knowledge gained while solving one problem and applying it to a different but related problem.

—wikipedia

Transfer Learning with Inception



Retrain Inception with custom dataset

Fork retrain script from TF repo

https://github.com/tensorflow/hub/blob/master/examples/image_retraining/retrain.py

Prepare Dataset

```
TF_DIR
/signature_dataset
    /None
        image_01.jpg …
/XXX
        image_02.jpg …
/Signature
        image_03.jpg …
```

Million Dollar Script

Run Retrain Script

```
$ export TF_DIR=/path/to/tf_stuff_dir
$ python retrain.py
    --bottleneck_dir=$TF_DIR/bottlenecks
    --how_many_training_steps 500
    --model_dir=$TF_DIR/inception
    --output_graph=$TF_DIR/graph.pb
    --output_labels=$TF_DIR/labels.txt
    --image_dir=$TF_DIR/signature_dataset
```

Retrain output

INFO:tensorflow **Step 0**: Train accuracy = 80.0%
INFO:tensorflow Step 0: Validation accuracy = 64.0% (N=100)
…
…
INFO:tensorflow **Step 499**: Train accuracy = 100.0%
INFO:tensorflow **Step 499**: Validation accuracy = **98.0%** (N=100)

```
TF_DIR
    /signature_dataset
    /inception
    /bottlenecks
    graph.pb
    labels.txt
```

Our Model

Our Labels (categories)

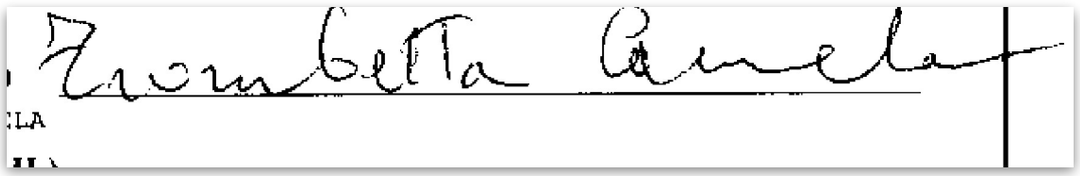
Predict an ORM class

```
def classify_images(image_path, label_path, graph_path):
    label_lines = [line.rstrip() for line in tf.gfile.GFile(label_path)]

    with tf.gfile.FastGFile(graph_path, 'rb') as f:
        graph_def = tf.GraphDef()
        graph_def.ParseFromString(f.read())
        tf.import_graph_def(graph_def, name='')

    with tf.Session() as sess:
        image_data = tf.gfile.FastGFile(image_path, 'rb').read()
        softmax_tensor = sess.graph.get_tensor_by_name('final_result:0')
        predictions = sess.run(softmax_tensor, {'DecodeJpeg/contents:0':
image_data})
        top_k = predictions[0].argsort()[-len(predictions[0]):][::-1]
        for node_id in top_k:
            label = label_lines[node_id]
            score = predictions[0][node_id]
    tf.reset_default_graph()
    return label, score
```

Results



Score: 0.99711 **Label: signature**
Score: 0.00183 Label: xxx
Score: 0.00106 Label: none

Score: 0.97478 **Label: xxx**
Score: 0.01302 Label: signature
Score: 0.01221 Label: none

Score: 0.98991 **Label: none**
Score: 0.00947 Label: signature
Score: 0.00063 Label: xxx

* ~ 300 ms to classify each image

Results

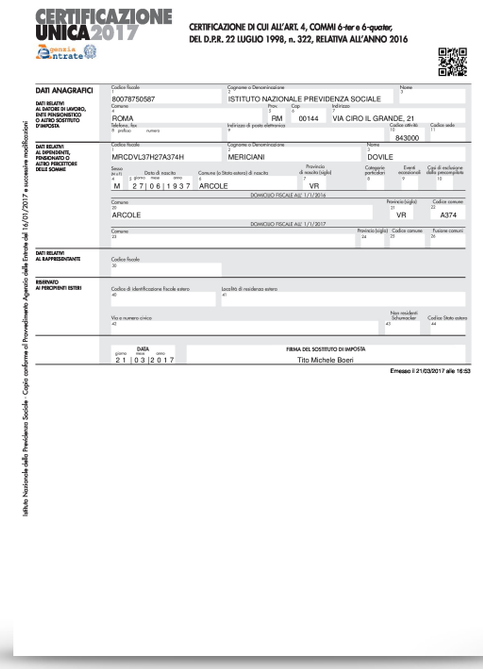
Before	After
Spaghetti code C + Python	Only Simple Python
WTF debug	Highly maintainable
Static decision tree	Smart decision tree
1-2 Developer in 3 months	1 Developer in 1 month
License $$$ K + Cloud costs	Less than $100 monthly cost

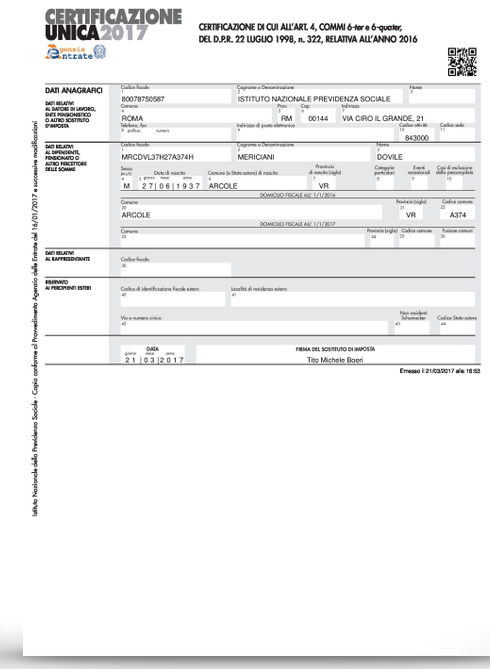
A.L.I Rule

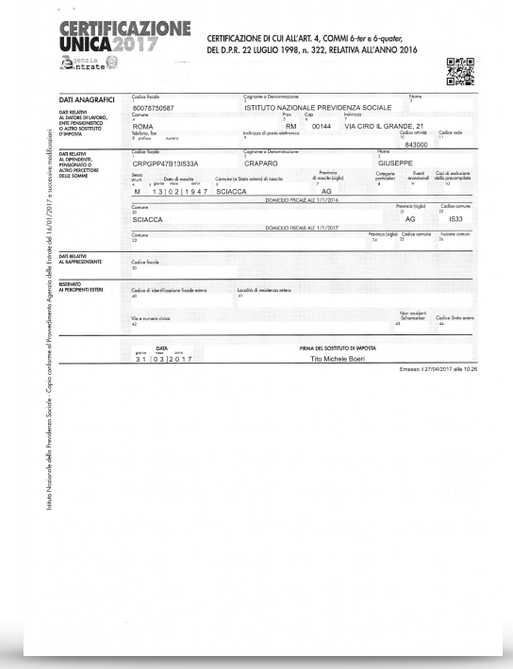
If you can identify a pattern in an image, then Inception can do it as well.

*A.L.I. Artificial Learning Intelligence :-)

Testing A.L.I. Rule







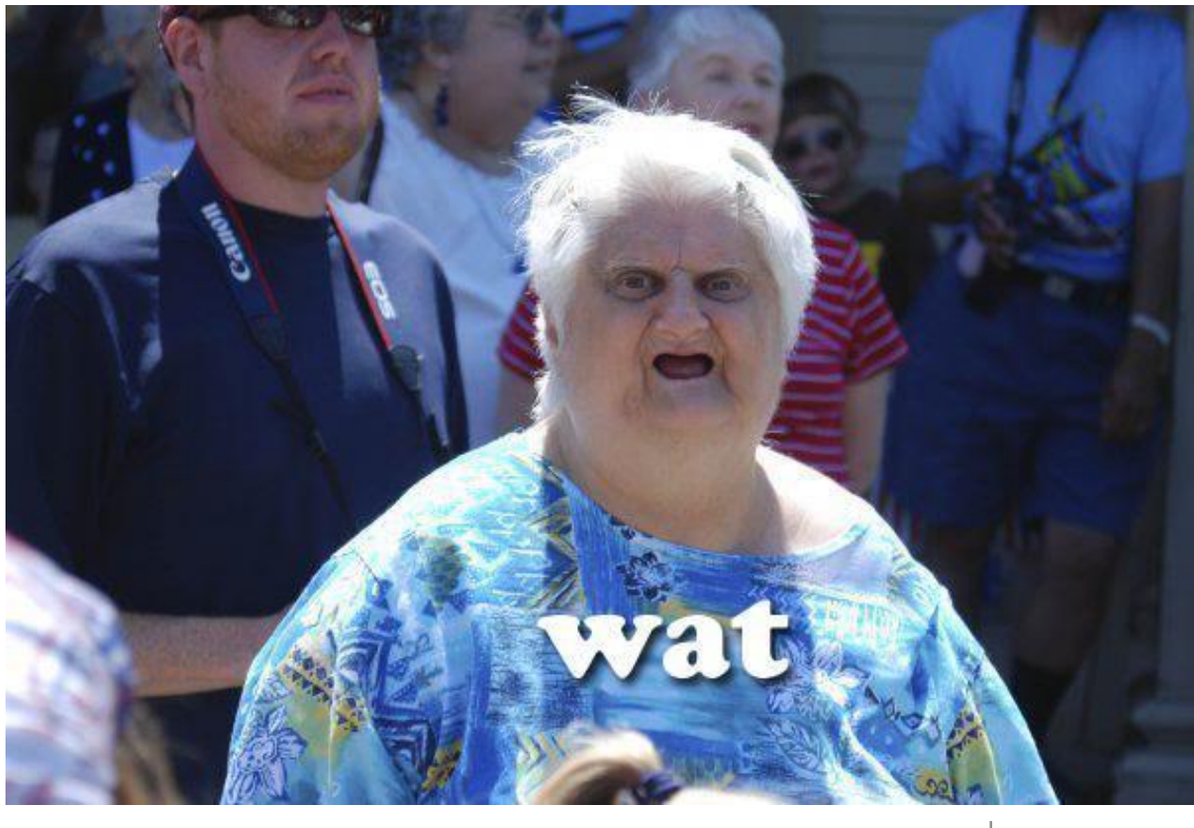
Out of defined domain documents

Predict this image using tax document classifier:

Out of defined domain documents

Predict this image using 730 document classifier:

Score: 0.81461
Label: CUD



Make Some Noise

Add some noise to our dataset using a dataset of pictures of (irrelevant) objects belonging to 101 categories.

http://www.vision.caltech.edu/Image_Datasets/Caltech101/

Documents will be distinguished from random images

Textual Classifier

Insurance receipts are not visually similar but they have many **common words (features)** in their text

La tua Assicurazione svizzera

helvetia

AGENZIA Novara A (178)

Egr. Sig.
BARONI VITTORIO
Via della Pace 3
20087 Robecco s/N MI

DICHIARAZIONE

PREMI DETRAIBILI

Polizza n. 178/08/1305

L'Helvetia Compagnia Svizzera d'Assicurazioni SA dichiara che:

il Sig. **BARONI VITTORIO** in base alle disposizioni fiscali (decreto legislativo n° 47/2000 del 18/02/2000 e successive modifiche ed integrazioni), ha corrisposto nel **2014** il seguente premio lordo Euro **439,00**

di cui detraibili per le singole garanzie:

Morte	Euro	**100,74**
Invalidità Permanente	Euro	**219,53**
Invalidità Permanente da malattia	Euro	**===**
Totale premio detraibile lordo	Euro	**320,27**

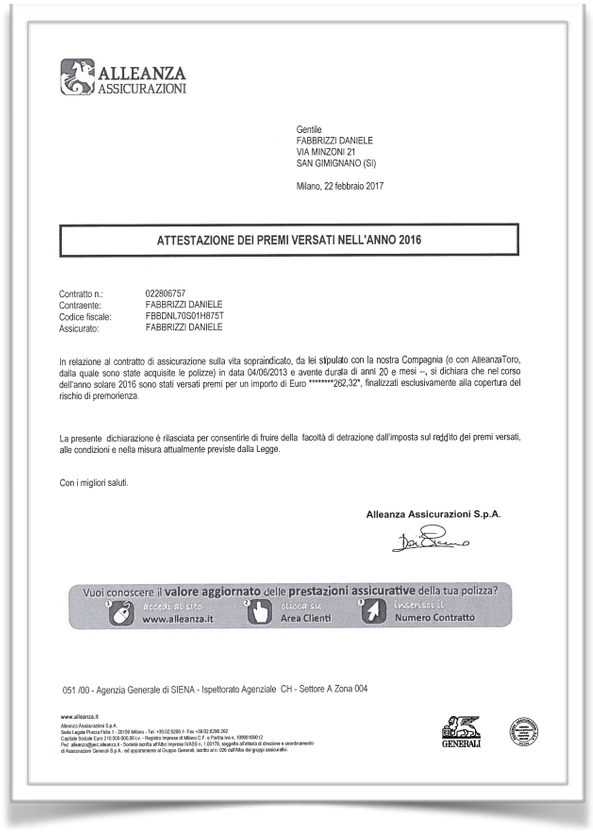
La presente dichiarazione viene rilasciata, per gli usi consentiti dalla legge, su richiesta del Contraente, che ha l'onere di verificare gli eventuali limiti previsti dalla citata normativa.

Milano, 30/04/2015

Helvetia Compagnia Svizzera d'Assicurazioni SA
Rappresentanza Generale e Direzione per l'Italia

Mod.AS – Ed. 03/2012

Helvetia Compagnia Svizzera d'Assicurazioni SA
Rappresentanza Generale e Direzione per l'Italia



ALLEANZA ASSICURAZIONI

Gentile
FABBRIZZI DANIELE
VIA MINZONI 21
SAN GIMIGNANO (SI)

Milano, 22 febbraio 2017

ATTESTAZIONE DEI PREMI VERSATI NELL'ANNO 2016

Contratto n.: 022806757
Contraente: FABBRIZZI DANIELE
Codice fiscale: FBBDNL70S01H875T
Assicurato: FABBRIZZI DANIELE

In relazione al contratto di assicurazione sulla vita sopraindicato, da lei stipulato con la nostra Compagnia (o con AlleanzaToro, dalla quale sono state acquisite le polizze) in data 04/06/2013 e avente durata di anni 20 e mesi --, si dichiara che nel corso dell'anno solare 2016 sono stati versati premi per un importo di Euro ********262,32*, finalizzati esclusivamente alla copertura del rischio di premorienza.

La presente dichiarazione è rilasciata per consentirle di fruire della facoltà di detrazione dall'imposta sul reddito dei premi versati, alle condizioni e nella misura attualmente previste dalla Legge.

Con i migliori saluti.

Alleanza Assicurazioni S.p.A.

Vuoi conoscere il **valore aggiornato** delle **prestazioni assicurative** della tua polizza?
1) accedi al sito www.alleanza.it 2) clicca su Area Clienti 3) inserisci il Numero Contratto

051 /00 - Agenzia Generale di SIENA - Ispettorato Agenziale CH - Settore A Zona 004

www.alleanza.it

ITAS VITA SPA

A063/ BOLZANO SUD

LARCHER SONJA
VIA ORTLES 11
39100 - BOLZANO (BZ)

Trento, 07/04/2017

Comunicazione dei premi di assicurazione sulla vita detraibili dall'imposta sui redditi delle persone fisiche
Polizza n. 704585/0001
Decorrenza 10/10/2016
Codice fiscale contraente LRCSNJ82C42A952J
Assicurato LARCHER SONJA

A seguito di quanto pagato nel corso del **2016**, le comunichiamo che il premio detraibile dall'imposta sul reddito delle persone fisiche ammonta ad **euro 94,00**, eventualmente da ricondurre all'interno del tetto di detraibilità valido per lo specifico anno di imposta.
Tale importo potrebbe non coincidere con il premio da lei versato in quanto, in base all'art. 15, comma 1, lett. f), del D.P.R. del 22 dicembre 1986, n. 917 (Testo unico delle imposte sui redditi), è consentita la sola detrazione della quota di premio afferente il caso morte.
La presente dichiarazione è rilasciata al fine di agevolare il contribuente in sede di compilazione della dichiarazione dei redditi.
Voglia gradire i migliori saluti.

ITAS VITA
Il Direttore Generale

*Abbiamo trasferito la nostra sede operativa e legale in **piazza delle Donne Lavoratrici 2, 38123 Trento.** I nostri rapporti contrattuali **non** subiscono alcuna variazione. Rimangono invariati denominazione sociale, dati fiscali, recapiti telefonici, indirizzi e-mail e numeri fax.*

VITA post 01.01.2001 – ed. 2014

Preparing dataset

A CSV with all of our features

	TEXT
doc 1	Extracted text by OCR
doc_2	Extracted text by OCR
doc_2	Extracted text by OCR

Preparing dataset

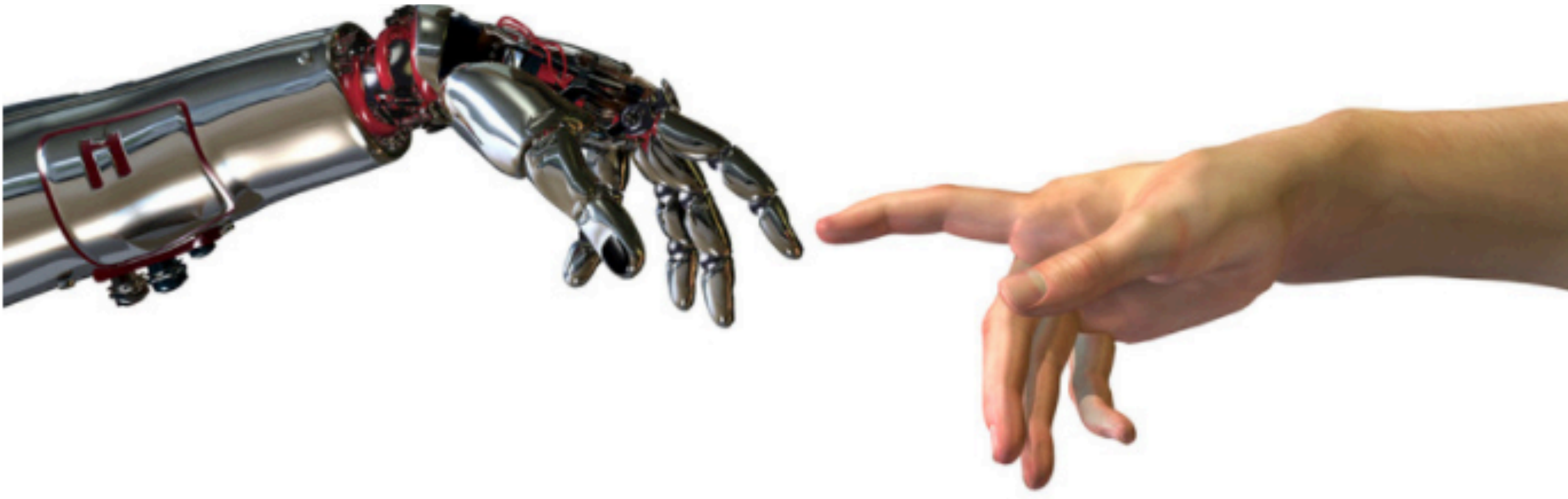
A **special categorical column** our target

	TEXT	Category
doc 1	Extracted text by OCR	Insurance
doc_2	Extracted text by OCR	Insurance
doc_2	Extracted text by OCR	ID

Preparing dataset

Provide more info about our data

	TEXT	Category	Common words	Insurance Common Words	ID Common Words
doc 1	Extracted text by OCR	Insurance	Insurance, Contract, Recipet	1	0
doc_2	Extracted text by OCR	Insurance	Life Insurance, Zurich,	1	0
doc_2	Extracted text by OCR	ID	Nationality, First Name, Surname,	0	1



AWS Machine Learning

or data science with one hand

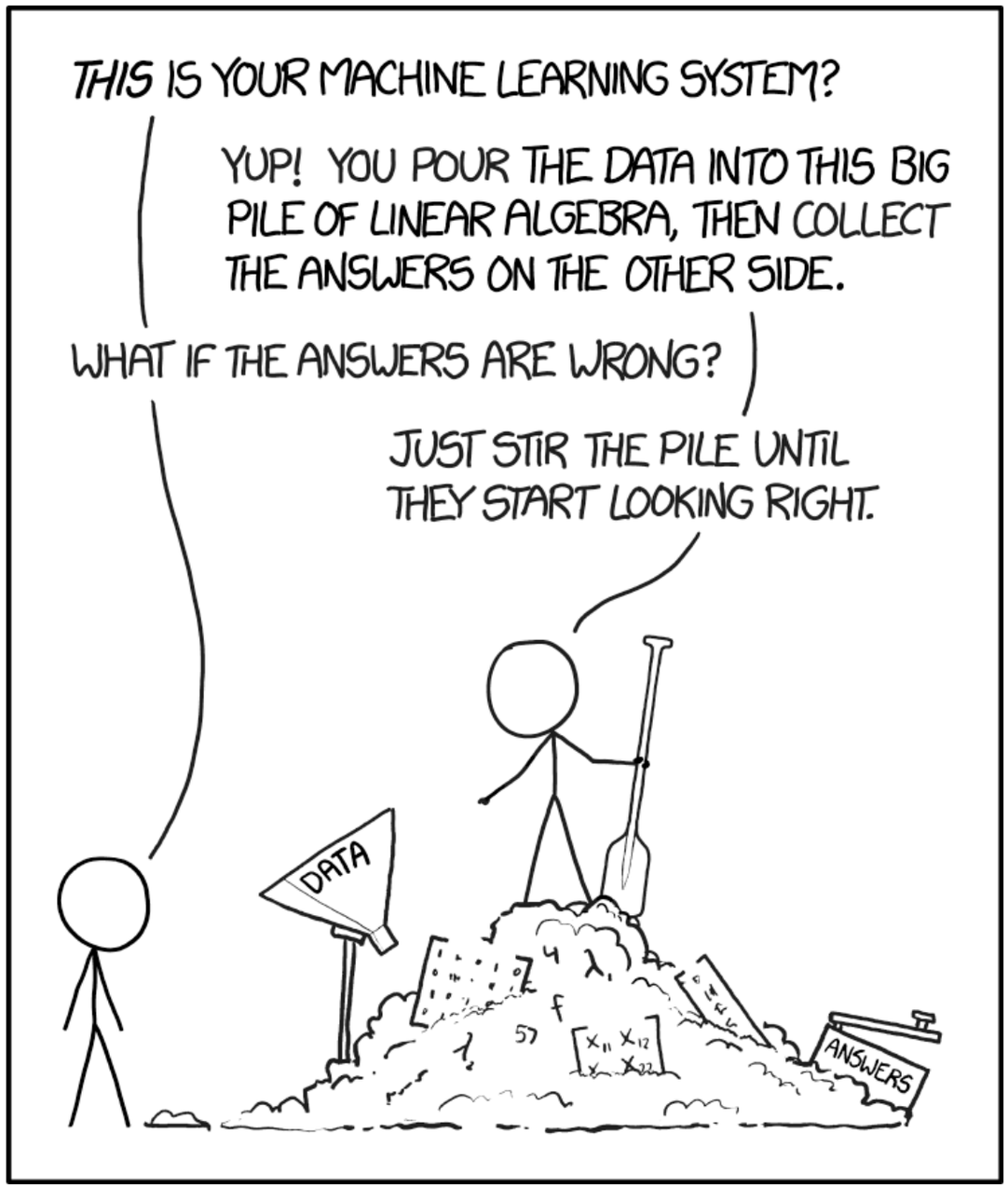
AWS ML: Datasource

Object (our CSV) that contains information about your input data

attribute names and types

(Possible values: Binary, Categorical, Numeric, Text)

Garbage in, Garbage out



Garbage in, Garbage out

Data must be as clean and consistent as possible

NY ≠ ny ≠ New York ≠ new_york

Garbage in, Garbage out

Data must be as clean and consistent as possible

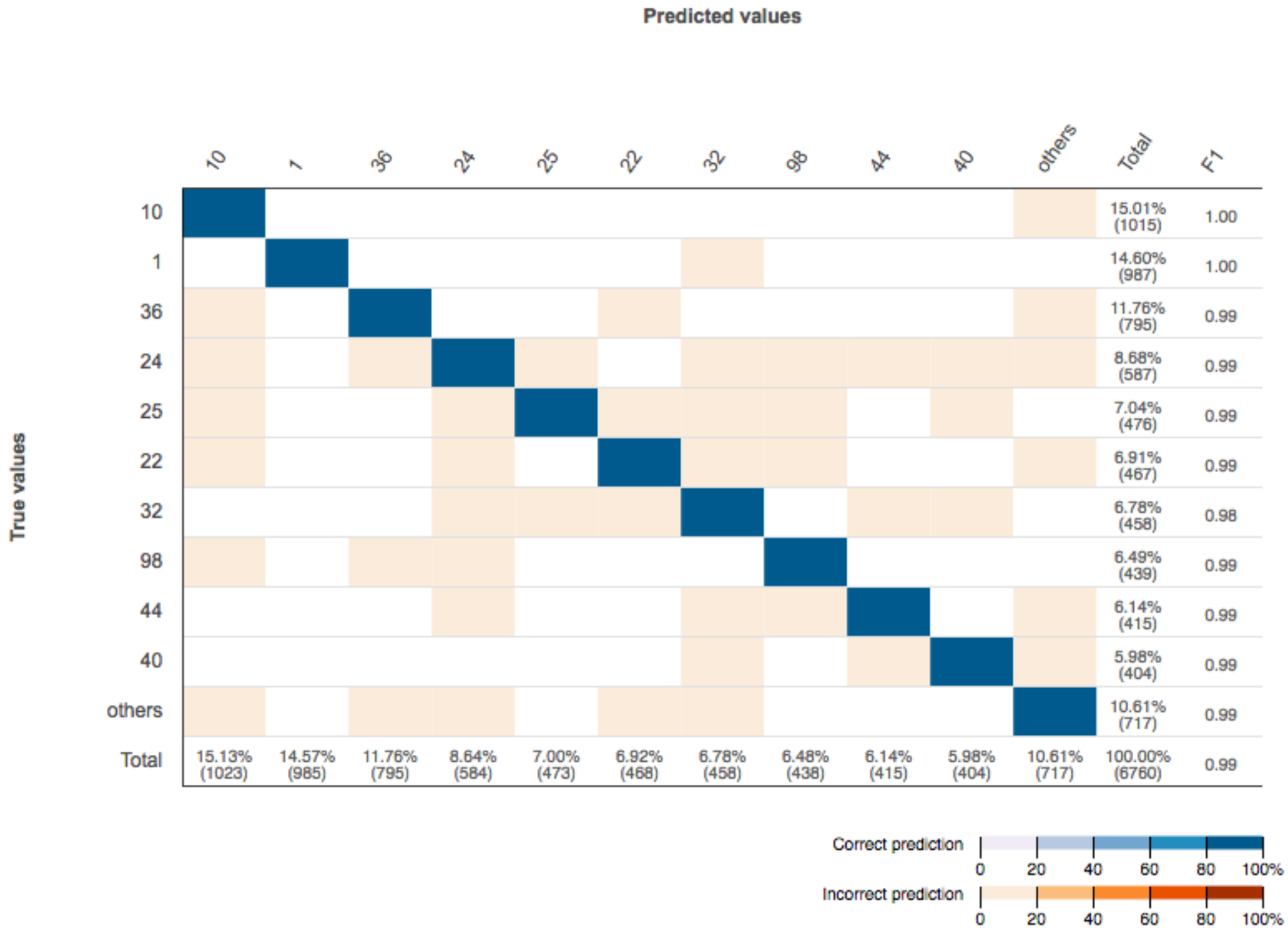

Wees niet langer te dik Wilde ook wel graag k
të\/Xin badpak, maar .... *^7#Kr_i> //
/fw/7*.**7 brulkert aetarOCt oaei ' _□?> //
APV '-/ r Lüoor ««««.Kor. f* oebrnxk*. «>«« l
(tf*-v\ \ \\K& -'-: ⌈TJr\W * MvA V / / I Ik
had er ef» hekel <□<•'• ••" ,bod. /'/V^ï V?ï,
*7 /-\ / BonKora .nu N«' I -M □•Hsf in / \v _
<i/ - r*.'**."*\ I \\ W^ "mi..... ... h. n-.mit.it.
/^••'•\rJiil__?\ "' Een VeiHa Verma'
Ar:.^'lt^S(_S_i£ geringsmiddel in ll^ir^C^ vloe-
ibaren vorm. / \r^ "*__£_£"__*.____ *& / / \
\ aanzemngei- rn voolil i» » .-k... en ' I l n.l! geen
i ong. \\\ _____»üt_ _- K« kunt flink eten von
«makeluk vo*'"J' * tt i*^r*^^\Y.l_. i" di bU elke
'--il □°"flor" *l'rt» "^ aangeraden. WaanM
e;.,fdé? af gewicht terug heM. neem dan nie
»•'«/□' en behoudt Uw nieuw, a.anke
«rjur.fc-K.»» bevat «eer apothekers en erkende
drogisten a f. 3.25 per flacon

AWS ML: Model

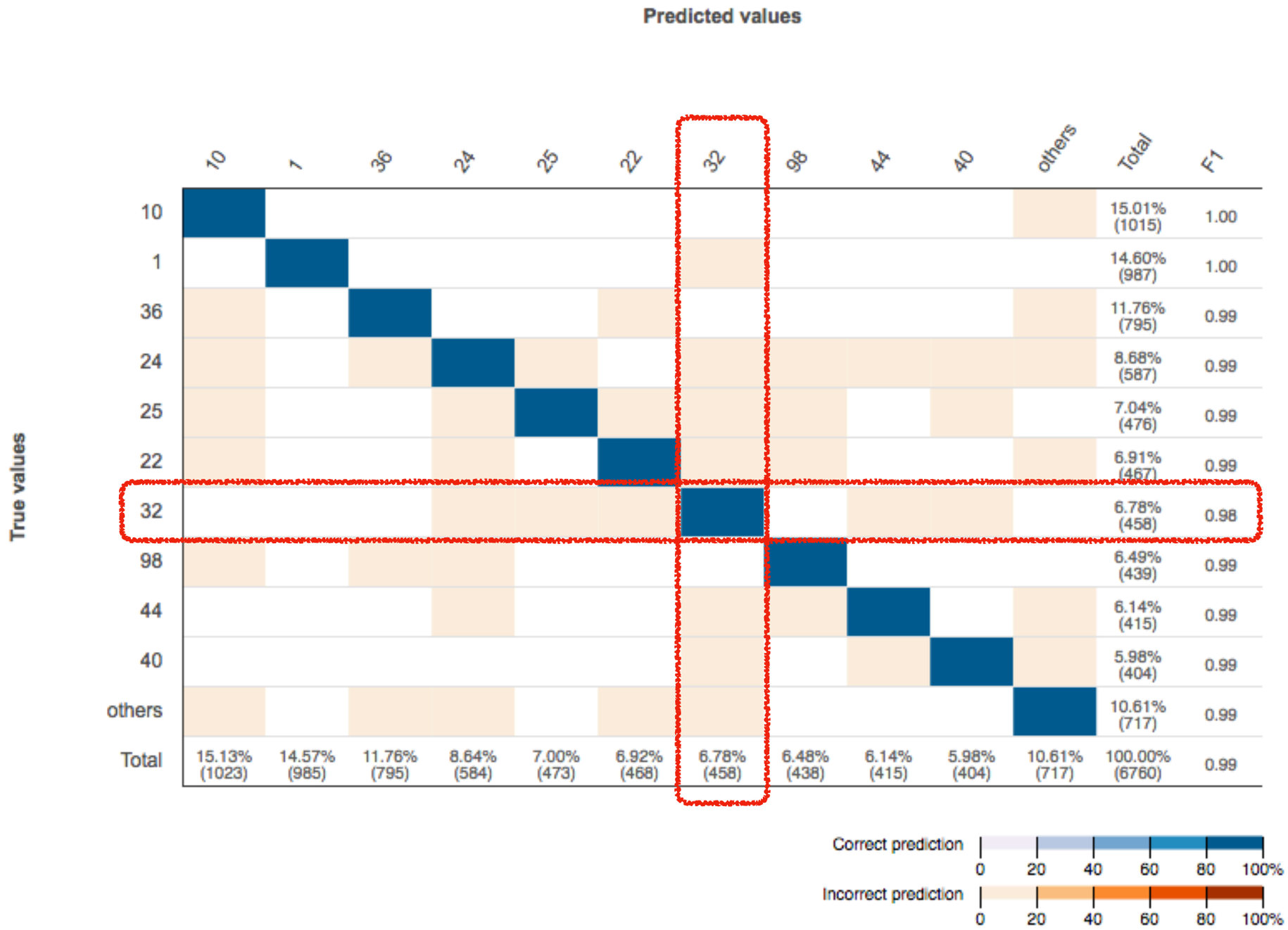
Don’t forget to shuffle your data. AWS ML will do it for you if didn’t do it during creation dataset.

When Evaluation is done a **Confusion Matrix** will be generated.

Confusion matrix



Confusion matrix



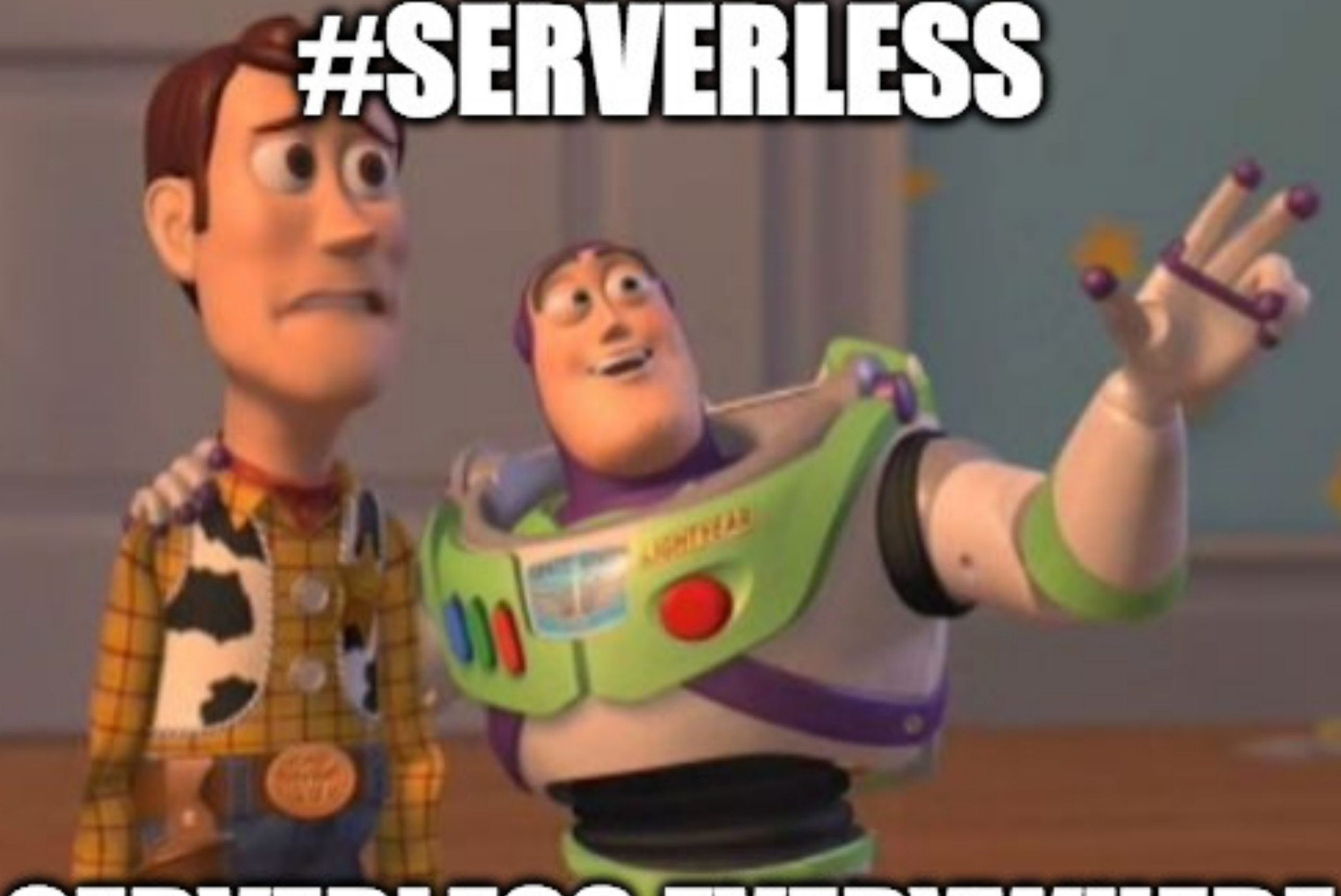
Predictions

Enabling the ML model for real-time predictions or using batch predictions.

Predict document category

```
def predict_document_category(ml_client, text, **columns_data):
    model_id = 'ml-5YzhSO3yt5u'
    endpoint= 'https://realtime.machinelearning.eu-west-1.amazonaws.com'
    record_flds = dict(text_data=text)
    record_flds.update(columns_data)
    try:
        response = ml_client.predict(
            MLModelId=model_id,
            Record=record_flds,
            PredictEndpoint=endpoint
        )
    except PredictionError as exc:
        raise PredictionError(u'Something went wrong\n{}'.format(exc))
    else:
        label = response['Prediction']['predictedLabel']
        score = float(response['Prediction']['predictedScores'][label])
        return label, score
```

#SERVERLESS



SERVERLESS EVERYWHERE

Shopping list to run a function in cloud

- VPC (Virtual Private Cloud)
- Load Balancer
- Server(s)
- Operating System
- Web Server
- (WSGI App Server in case of Python)
- Tuning programming environment

FaaS (Function as a Service)

Run code without thinking about servers

FaaS (with AWS Lambda)

Lambda should be **triggered.**

Possible triggers

AWS services (Handling Upload file on S3 event)

HTTP Endpoints (Handling Calls on API Gateway event)

API Gateway

Run API's without ~~servers~~ **backend**

Lambda Handler

```
def lambda_handler(event, context):
    job_id = uuid.uuid4().hex
    response = {
        'job_id': job_id,
        'creation_date': datetime.datetime.now().isoformat()
    }
    query_string = event['params']['querystring']
    svc = query_string['svc']
    data_bytes = base64.b64decode(event.get('body-json'))
    put_object_to_s3(data_bytes, svc, job_id)

    create_job(svc=svc, key_id=job_id)

    celery_app.send_task(
        'my_task_name',
        args=(svc, job_id)
    )
    return response
```

Going serverless - PROS

No infrastructure

Going serverless - PROS

Python, Node.js, Java, C#, Go

Going serverless - PROS

Tightly integrated with other AWS services

Going serverless - PROS

Scalable to (almost) ∞

Going serverless - CONS

Serious limits imposed by the vendor

Going serverless - CONS

Spin-up time

Going serverless - CONS

Less than a server

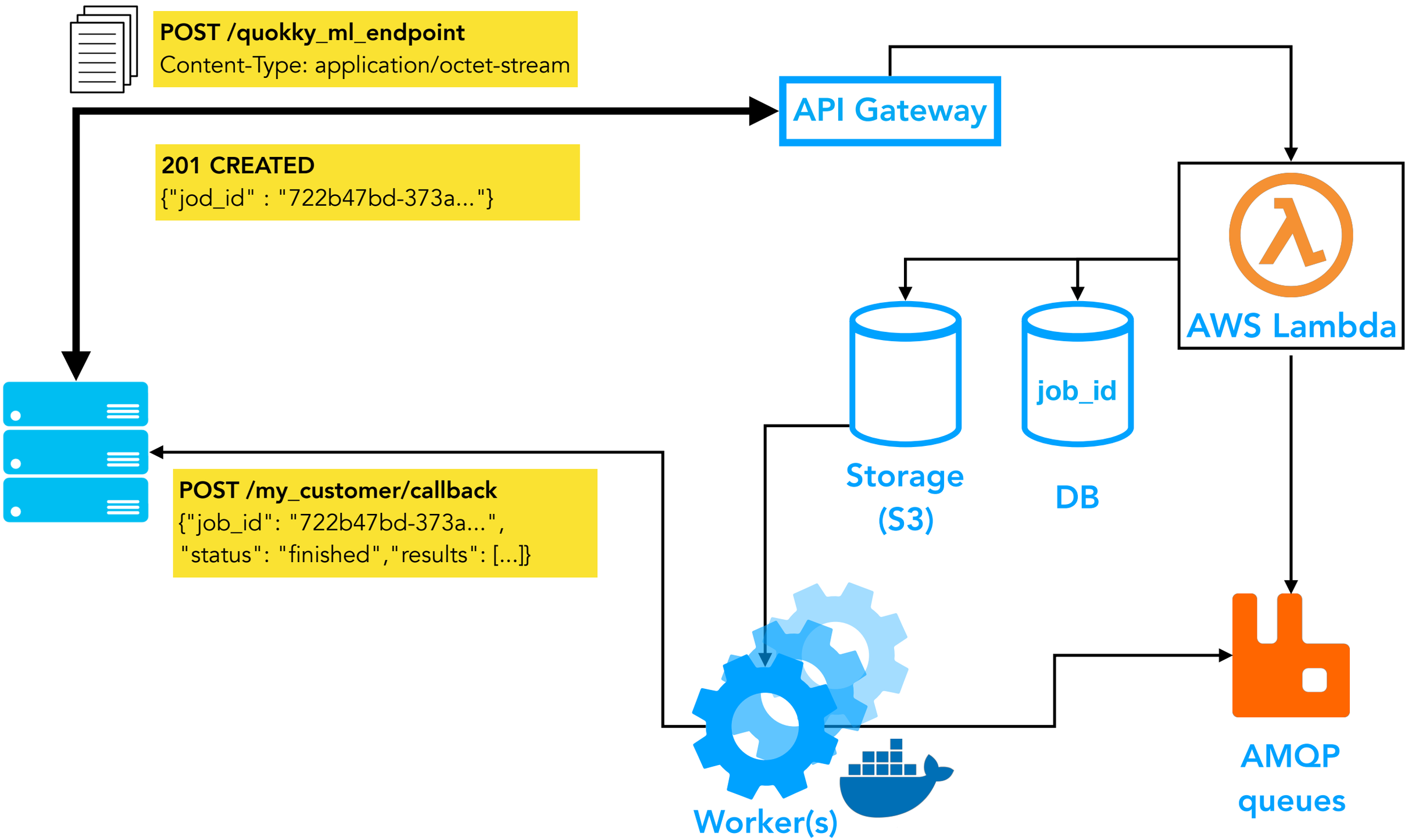
Going serverless - CONS

Could become expensive

Put everything together

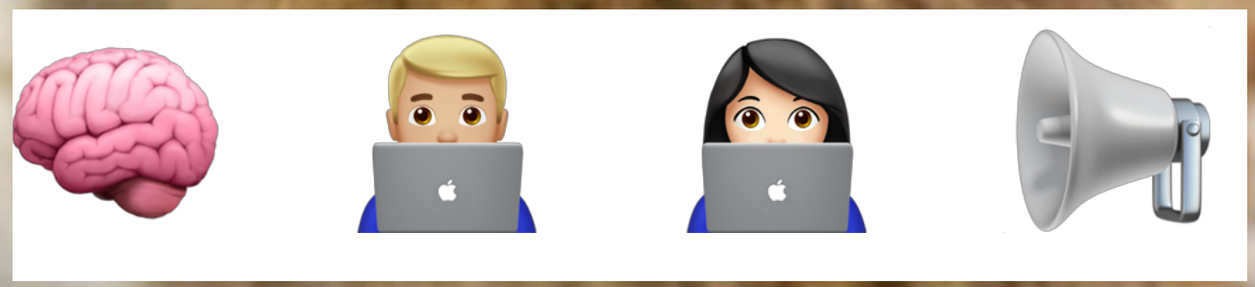
AWS API Gateway
AWS Lambda
AWS S3
MongoDB
RabbitMQ
AWS EC2 Container Service (to run ML logic)

Going serverless



QUOKKY IS HIRING!

jobs@quokky.com



Thank you Brandon

All images for CNN part of this talk have been taken from this great presentation on CNN's by Brandon Rohrer.

You may find the presentation here.

http://brohrer.github.io/how_convolutional_neural_networks_work.html


BEATLES