Facial Recognition

Paul Hrycewicz
Computer Science
October 20, 2020

Conclusion

- There is no magic, nor true understanding, by machines or software algorithms.
- There may be the *appearance* of understanding, or intelligence, by advanced technology or by sufficiently clever algorithms.
- Two parts to image recognition:
  - (1) Finding and identifying the image in a picture
  - (2) Using the image's features in a predictive model

Agenda

- Predictive modeling overview
- Image Representation
- Feature Recognition
- Convolutional Neural Networks
- Facial Recognition
- Accuracy
- Face Databases

Predictive Modeling

- A predictive model describes a set of data in such a way that it can predict the values of one or more parameters.
- All modeling is based on statistical analysis. There is no "magic" involved. Modeling simply uses the patterns that exist in data to develop probabilities of events.
- It would be extraordinarily time-consuming ("analytically intractable") for a human to develop an accurate predictive model.
- Numerous modeling software packages exist at reasonable cost from vendors like Google, SAS, Amazon, and Microsoft.

An Example of a Predictive Model

- Large trucking company has 2,000 trucks and 1,600 drivers.
- The driver churn rate is 70% - 7 out of 10 drivers leave the company each year.
- Training a new driver takes 3 months and costs the company $14,000.
- This represents a $15.7M annual expense for the company.
- Can we use a predictive model to predict which drivers are most likely to churn, so that the company can take proactive steps to prevent?
- If we can reduce the churn rate by 10%, that represents a $1.5M annual savings for the company.

Data for the Model

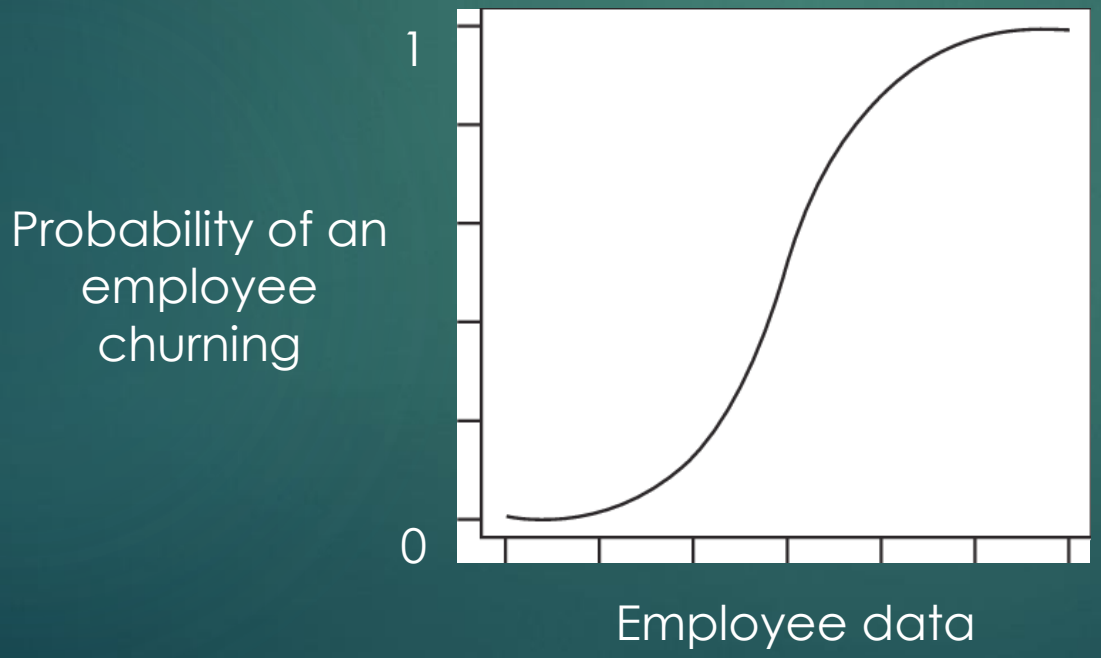
- Employee age
- Employee gender
- Employee home location
- Years driving
- Years with company
- Number of traffic tickets
- Family size
- Favorite brand of blue jeans

Sample Data

Name	Gender	Home zip	Years exp	Years empl	Tickets	Fam size	Blue Jeans	Did they churn?
Joe	M	95083	4	2	0	2	Levi's	No
Sam	M	15097	2	1	3	4	Wrang	Yes
Mary	F	32807	3	3	1	2	Levi's	No
Kim	F	10056	13	2	2	0	Wrang	Yes
Teddy	M	75234	8	5	3	8	OldN	Yes
Sally	F	90432	1	0	7	2	Gap	No
Davey	M	60604	4	2	3	4	Levi's	Yes

Logistic Regression Model

- This is only ONE method out of MANY that techniques for modeling.
- LR returns the probability of an event happening – a number between 0 (low probability) and 1 (high probability)
- Expressed as a probability curve, which is closely related to the bell curve



Building and Using the Model

1. Supervised Learning - with a training set

Age	Gender	Home zip	Tickets	…	Churn
37	M	95053	2		N
23	F	60603	7		Y
42	D	32788	0		N

Develop Model

Model

2. Scoring – determine the answer for a particular piece of data

Data on employee	
. . .	

Churn?
.832754

Image Recognition

Images of Queen Elizabeth

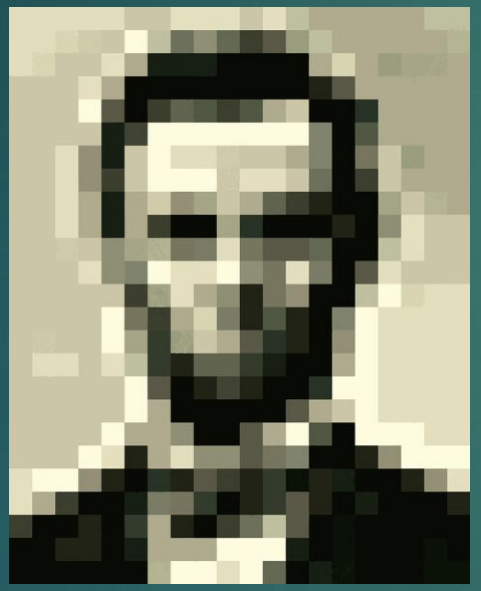
2002 (age 76)

2020 (age 94)

2020 (age 94)

Image Recognition

- An image is recognizable even if it contains only a small amount of data.



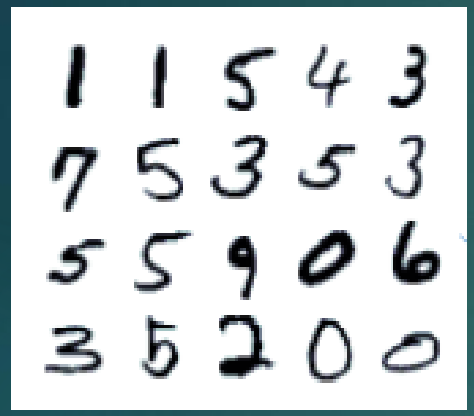
20x25 = 500 bytes

22x21x3 = 1,386 bytes

- A good resolution color picture on a web page is typically about 250K (but 25K when compressed, typically at 10:1). An HDTV frame is 18.6 MB.
- If a human can recognize the image, an algorithm should be able to as well.

Image Recognition – Digit Classifier

Training Set



1	1	5	4	3
7	5	3	5	3
5	5	9	0	6
3	5	2	0	0

Model

image

2

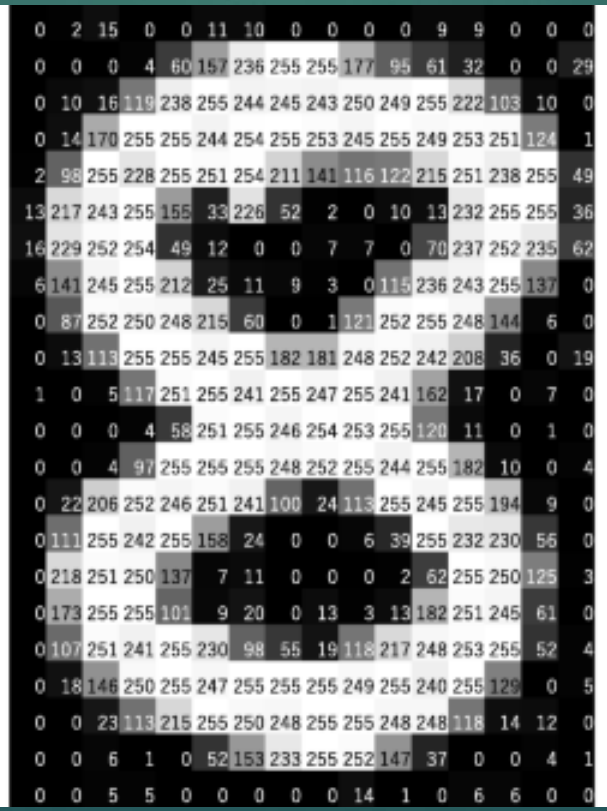
numeric value

But – what is  **?? It's not a variable with a numeric value like an age or a zip code.**

Image Representation

A picture is a set of pixels, each with a numeric value.

lighter = higher values
darker = lower values



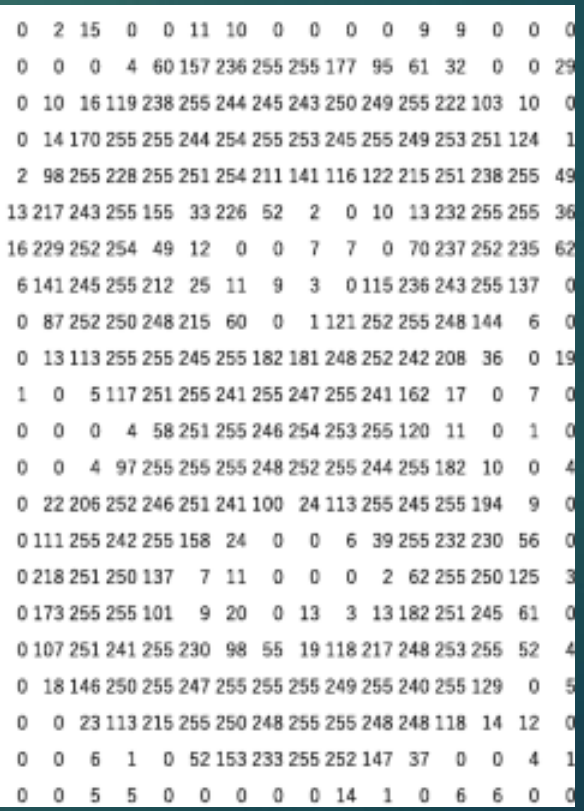
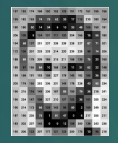
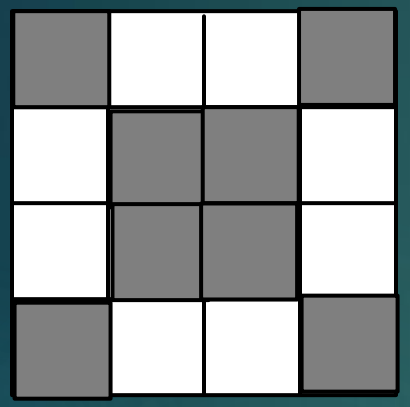


Image Representation



157	153	174	168	150	152	129	151	172	161	155	156
155	182	163	74	75	62	33	17	110	210	180	154
180	180	50	14	34	6	10	33	48	106	159	181
206	109	5	124	131	111	120	204	166	15	56	180
194	68	137	251	237	239	239	228	227	87	71	201
172	105	207	233	233	214	220	239	228	98	74	206
188	88	179	209	185	215	211	158	139	75	20	169
189	97	165	84	10	168	134	11	31	62	22	148
199	168	191	193	158	227	178	143	182	106	36	190
205	174	155	252	236	231	149	178	228	43	95	234
190	216	116	149	236	187	86	150	79	38	218	241
190	224	147	108	227	210	127	102	36	101	255	224
190	214	173	66	103	143	95	50	2	109	249	215
187	196	235	75	1	81	47	0	6	217	255	211
183	202	237	145	0	0	12	108	200	138	243	236
195	206	123	207	177	121	123	200	175	13	96	218

Looking for an "X"



picture of 'X'

0	1	1	0
1	0	0	1
1	0	0	1
0	1	1	0

pixels

remember
low = dark
high = light

1	0
0	1

\ filter

0	1
1	0

/ filter

activation map

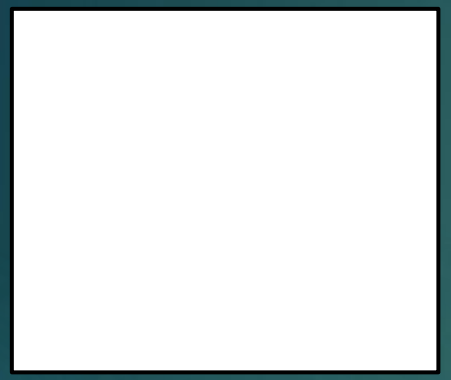
looking for \

looking for /

A convolution with two activation maps

Do We See an "X"?

Brute Force Approach



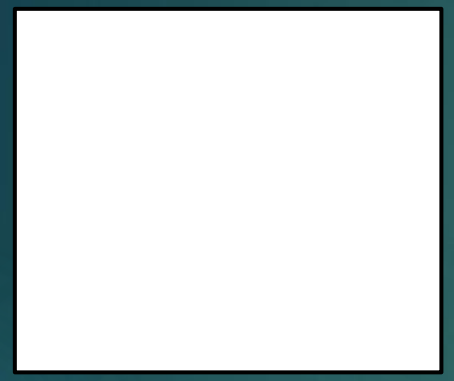
activation maps

Given the activation maps,
we can determine if an "X" is present or not.
Can use "brute force" approach:

IF the pattern in the "\" map is
three 0's in a diagonal
from top left to bottom right
AND
three 0's in a diagonal
from top right to bottom left

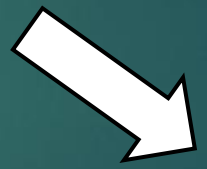
THEN it's an "X"

Do We See an "X"?
Supervised Learning Approach.



activation maps

flattened activation map



Model

A Better Approach…Feature Detection

- A feature is a part of an image that we consider significant.
- Low-level features include edges and curves.
- Higher-level features might be wheels, wings, or eyes.
- We use a convolutional ***filter*** to detect features.

- A filter enables the computer to "see."
- A filter transforms a set of pixels (numbers) into something that an algorithm can interpret.

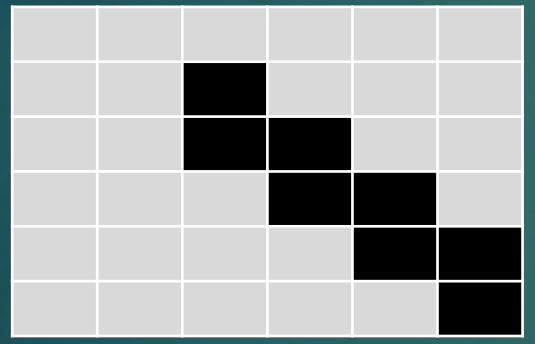
Example: A filter to detect a vertical edge. It "looks" for a dark line on the left of a light line.

-1	0	1
-1	0	1
-1	0	1

Detecting Features - \

Picture

1	1	1	1	1	1
1	1	0	1	1	1
1	1	0	0	1	1
1	1	1	0	0	1
1	1	1	1	0	0
1	1	1	1	1	0



remember
low = dark
high = light

Filter

1	0	0
0	1	0
0	0	1

filter to look for \

Convolution

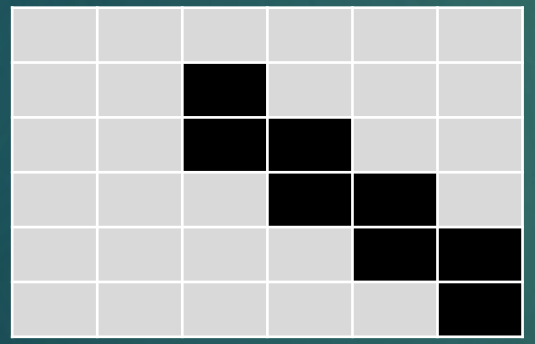
noise reduction:
ignore anything
>0? >1?

do we care about dark->light vs light->dark?
If we do, we can adjust the filter.
If we don't, we can use one filter and take the absolute value

Detecting Features - /

Picture

1	1	1	1	1	1
1	1	0	1	1	1
1	1	0	0	1	1
1	1	1	0	0	1
1	1	1	1	0	0
1	1	1	1	1	0



Filter

0	0	1
0	1	0
1	0	0

filter to look for /

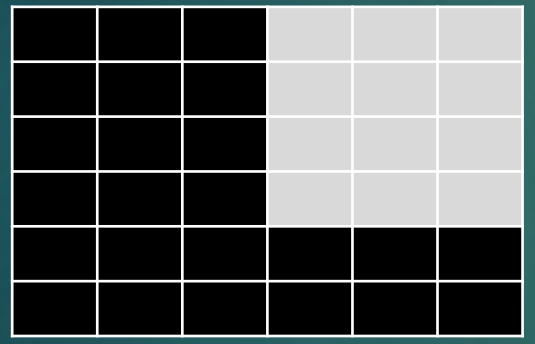
Convolution

The filter will ignore random noise.
Non-random noise may present a problem.

Detecting Features - Edges

Picture

0	**0**	**0**	**1**	**1**	**1**
0	0	0	1	1	1
0	0	0	1	1	1
0	0	0	1	1	1
0	0	0	0	0	0
0	0	0	0	0	0



Filter

-1	0	1
-1	0	1
-1	0	1

filter to look for dark->light vertical edge

Convolution

noise reduction: ignore anything < 3?

Curve filter

0	**0**	**0**	**0**	**1**	**0**
0	0	0	1	0	0
0	0	1	0	0	0
0	0	1	0	0	0
0	0	1	0	0	0
0	0	0	1	0	0

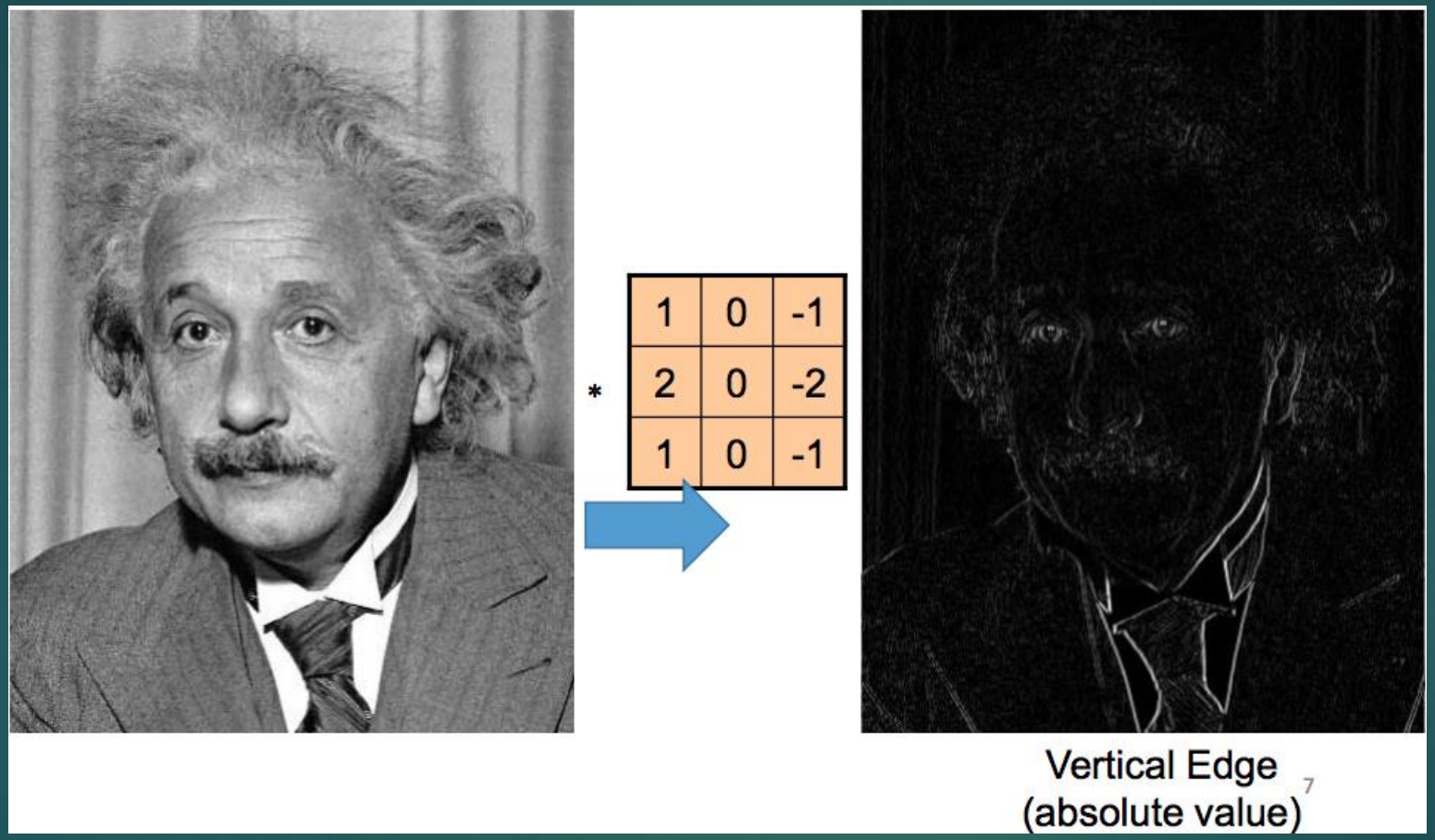
So…

- Why not design a filter for a face?

- We don't know where in the image the face is.
- We don't know how large the face is.
- All faces are different.
- We'd need many filters, and each would be huge.
- There's a big difference between recognizing something as a face in an image, and identifying that face as belonging to a specific person.

- Better to start with the basic features found in an image, and then determine if they build into a face.

Edge Detection on an entire image

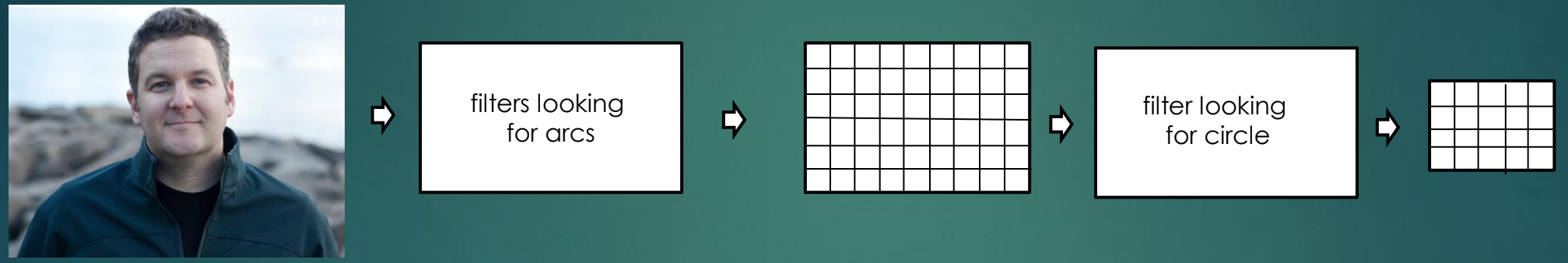


Sharpening filter

Easier to discern features from an edge

Pooling

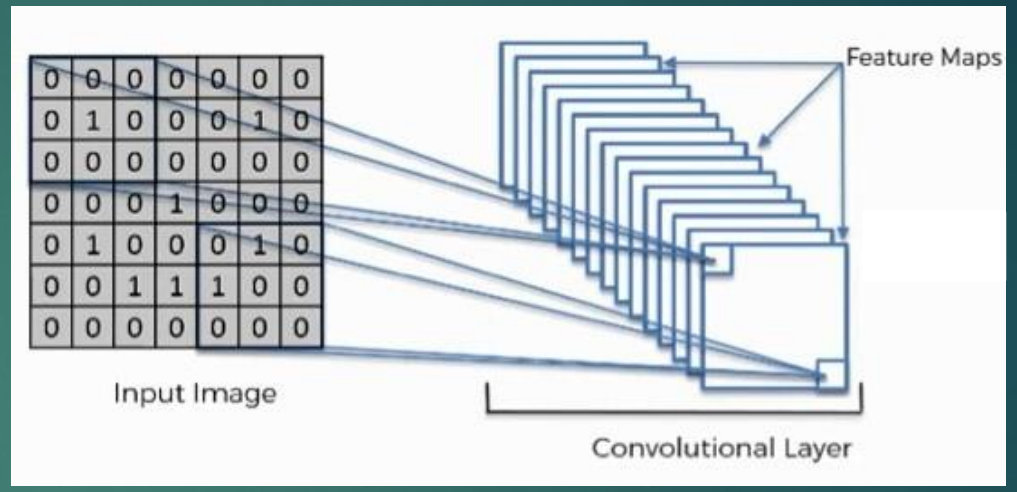
We can take the convolution and convolute it further to condense it.



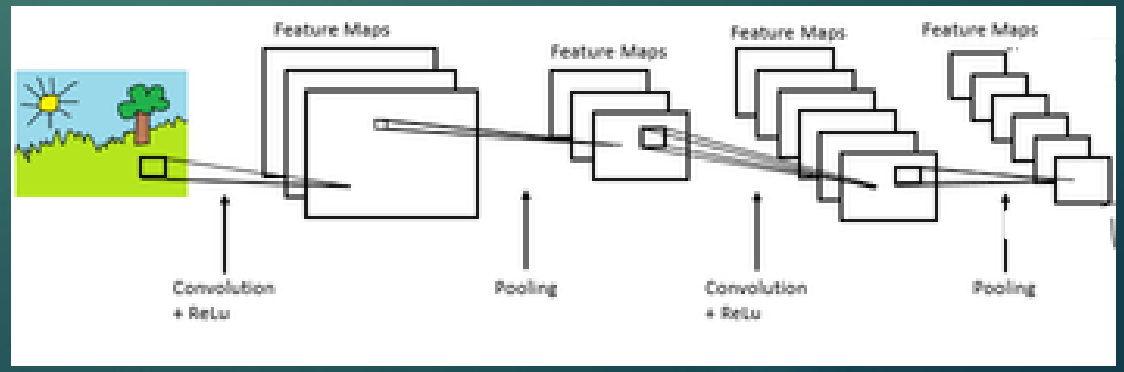
Feature Maps

- filters looking for arcs
- filters looking for horizontal half-ellipses
- filters looking for vertical half-ellipses
- filters looking for horizontal lines

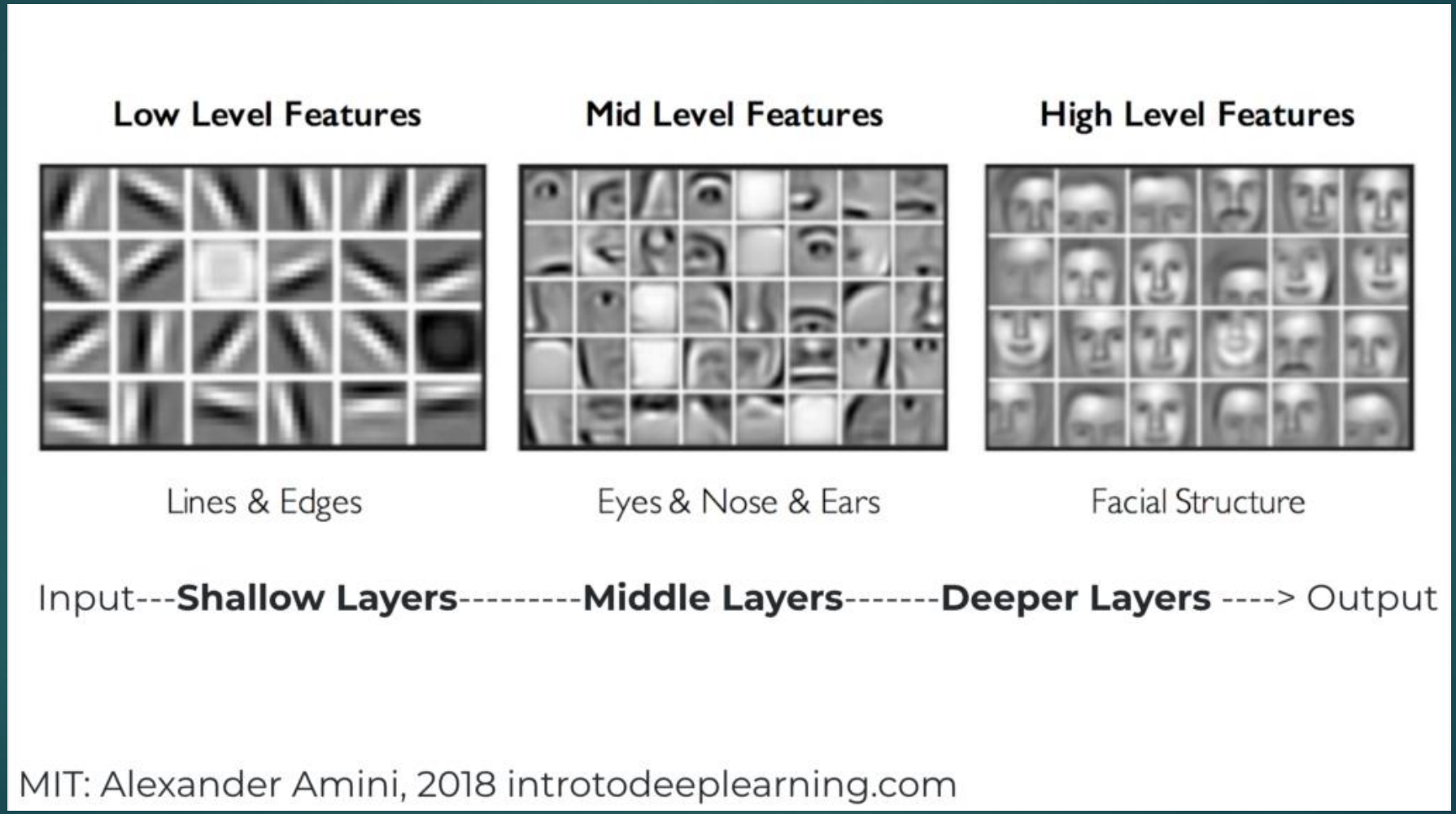
• • •



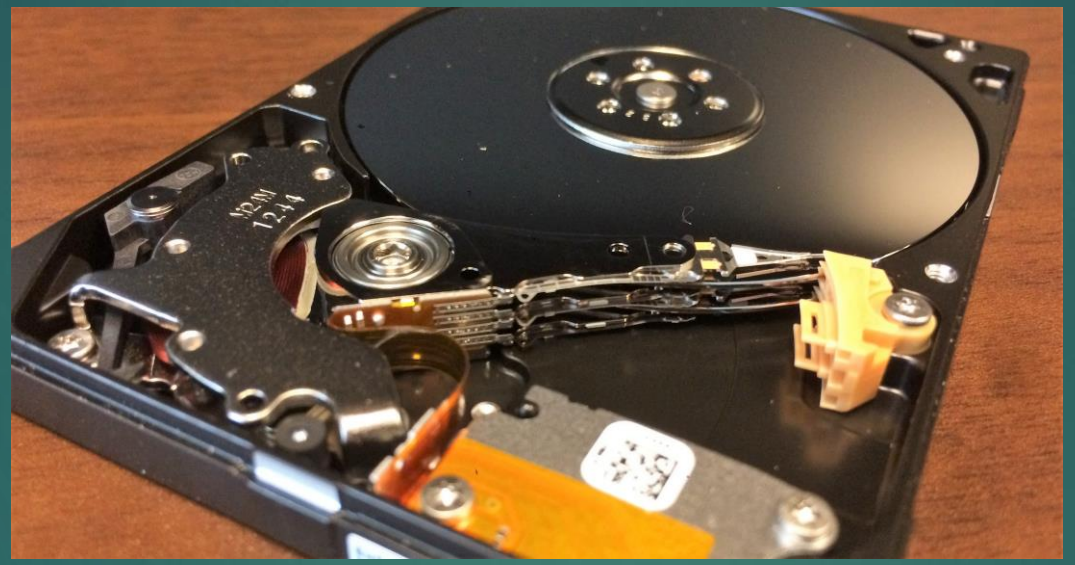
In this way, we get a condensed representation of the features present in an image.



Convolutional Layering

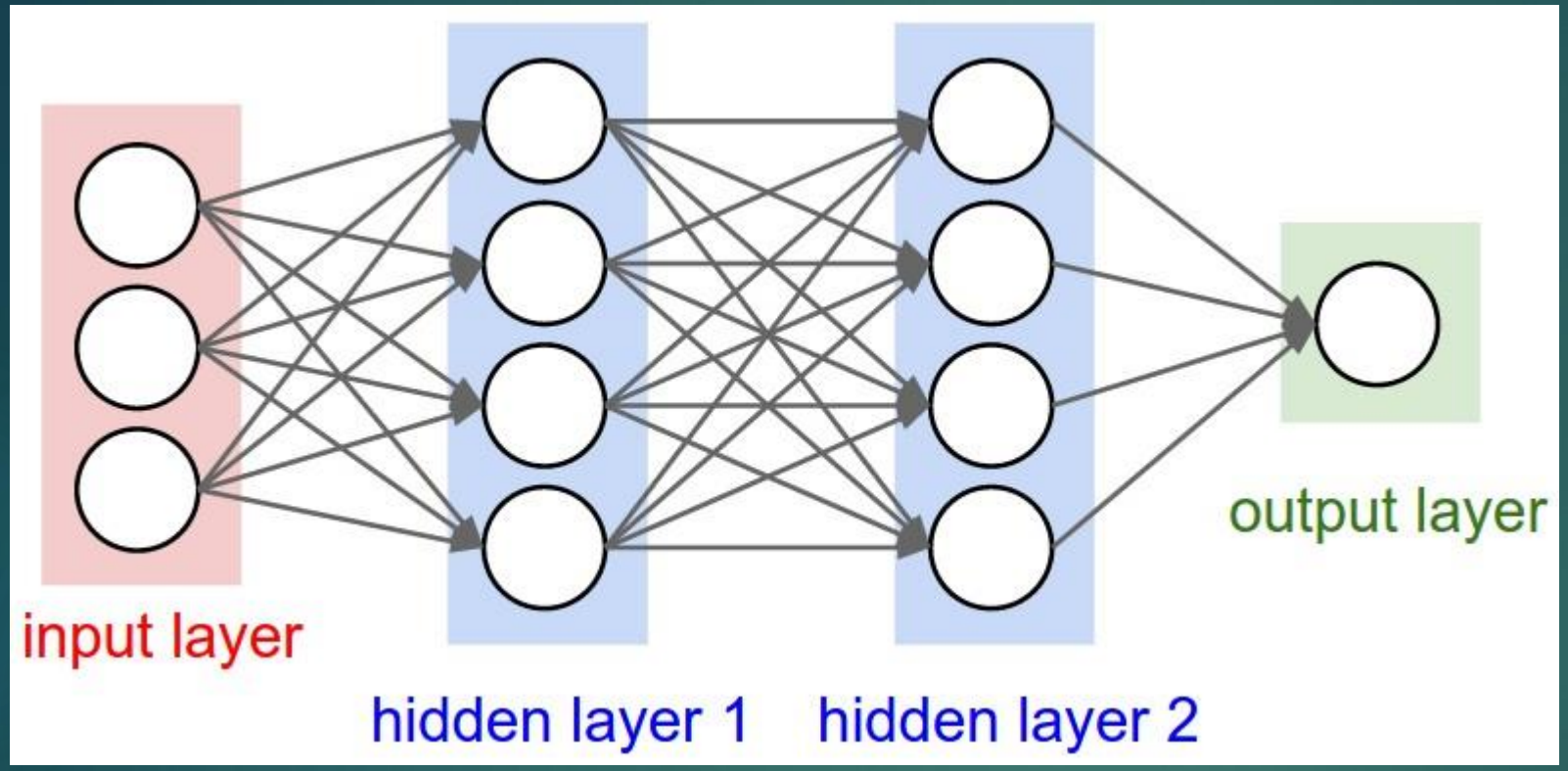


Commercial uses of general-purpose image recognition systems



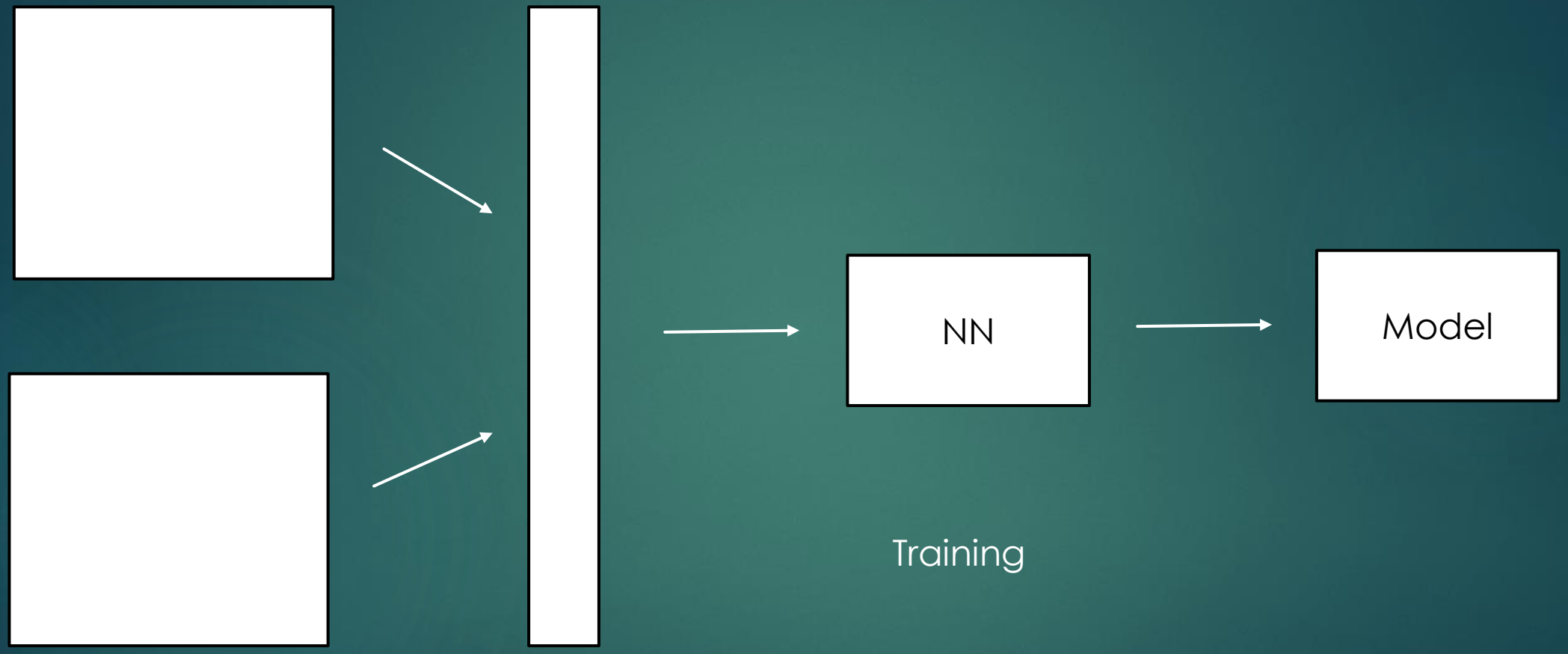
Microsoft, Google, and Amazon offer services which learn how to recognize objects, based on images and labels which you provide.

Artificial Neural Networks (ANNs)

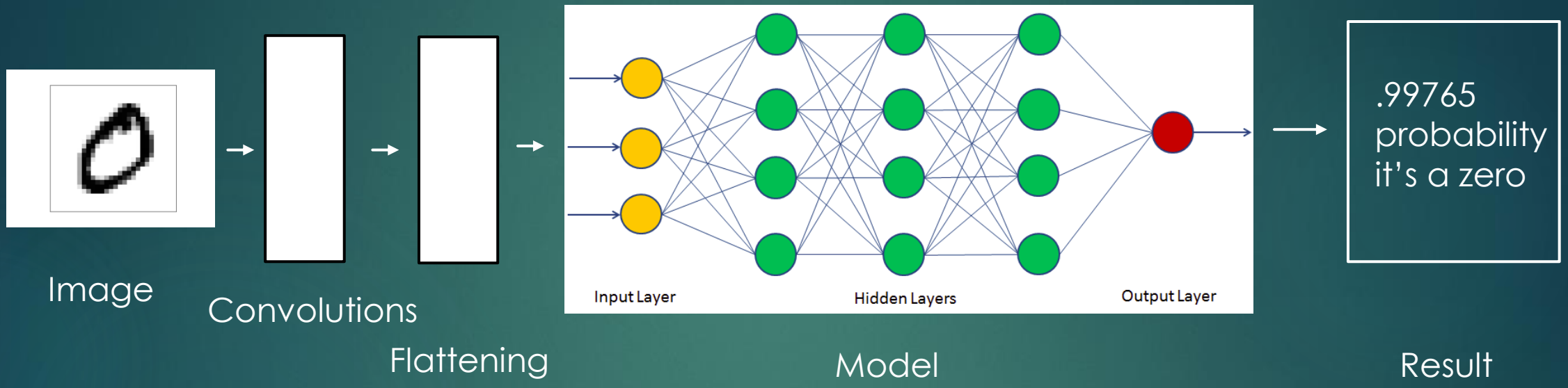


The NN transforms a set of inputs into an output.
This is what our brains do.
What was the name of the old movie with Jack Nicholson about a nurse and a hospital?

Using a NN for the Digit Classifier:
Step 1. Training the "fully connected" layer (the ANN)



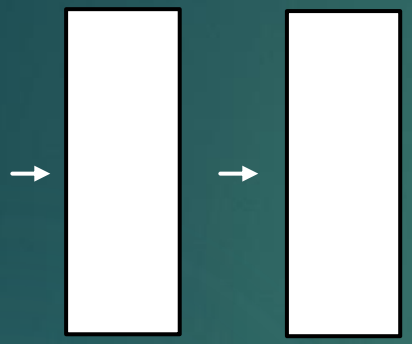
Using a NN for the Digit Classifier:
Step 2. Scoring



- This is called a convolutional neural network (CNN) – a set of convolving layers which develop inputs into a neural network.
- The fundamental theory behind a CNN is that pixels which are close together are more likely to be related than pixels far away from each other.
- This is how our brain recognizes what is seen by the eyes.

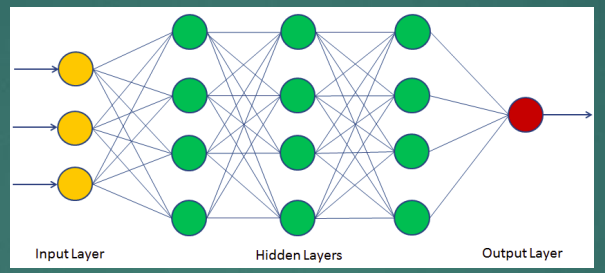
Facial Recognition

Image



Convolutions

Flattening



Model

Result

- Gender
- Age
- Facial Feature Map

Database lookup

Queen Elizabeth II
Buckingham Palace,
Westminster, London, SW1A 1AA
b. 1926, Mayfair, London

Facial Feature Map


```
"faceId": "5af35e84-ec20-4897-9795-8b3d4512a1f9",
    "faceRectangle": {"width": 60, "height": 60, "left": 276, "top": 43},
    "faceLandmarks":
     {"pupilLeft": {"x": "295.1","y": "56.8"},
      "pupilRight": {"x": "317.9","y": "59.6"},
      "noseTip": {"x": "311.6","y": "74.7"},
      "mouthLeft": {"x": "291.0","y": "86.3"},
      "mouthRight": {"x": "311.6","y": "88.6"},
      "eyebrowLeftOuter": {"x": "281.6","y": "50.1"}
```

How does commercial-grade FR work?

- Take stills from video camera feed
- Find faces
- Develop features for each face:
- gender, age, ethnicity, measurements
- Search facial image database

- The actual algorithms used by commercial companies' services are proprietary
- Companies do not supply the facial image database – the user does

Accuracy

- Generally measured in 9's – 99%, 99.9%, 99.99%, etc.

- There is an industry accuracy benchmark: The NIST FRVT (National Institute of Standards and technology Facial Recognition Vendor Test).

- False Positive – mistakenly identifies a face as belonging to a person
- False Negative – fails to recognize a face when it exists in the database

- 99.99% accuracy means that 9,999 out of 10,000 faces are recognized without error – therefore,
- 1 out of 10,000 is not and may be returned as a false positive or false negative.

- Commercial facial recognition services such as Amazon Rekognition and Microsoft CNTK are careful not to advertise any particular application for their technologies, nor guarantee any accuracies. It is generally thought that the accuracy is as high as 99.97% for these systems, given ideal conditions.

- In 2019, SFO had an average of 157,735 passengers per day. 99.97% accuracy means that there could be as many as 47 incorrectly identified passengers per day.

- When field-quality photos are used, such as those obtained from cameras scanning wide areas at airports, the error rate rose to as high as 9.3%. Subjects may not be looking directly into the camera, and may be partially obscured by shadows or other objects.

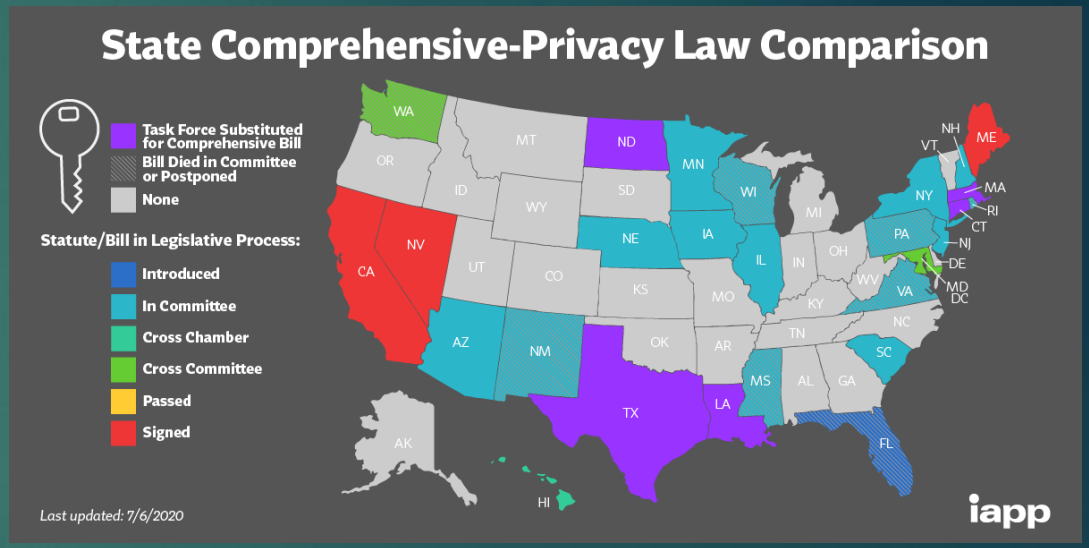
- Error rates increased as much as 10x when mugshots 10 years old were used.

Facial Image Databases

- Most images available on the web have not been privacy-protected.
- Some image databases are restricted; states' DL photos, for example.
- ICE, DOJ, and FBI are already using states' drivers licenses photos and mugshots.
- Inter-agency agreements allow law enforcement to share data.

- Only some data – PII and HIPAA – are federally protected.
- Some states have a Consumer Privacy Act (modeled after the CCPA) which allows you to opt out of data collection.
- Of course, people voluntarily submit data when the post to social media platforms.

- San Francisco has banned the use of facial recognition by its agencies.



Facial Image Databases

- Amazon, Microsoft, and Google do not maintain facial image databases. They require you to maintain your own.
- Clearview AI is an example of a company that markets both a facial recognition service and an underlying database of faces and identifying information. Its database is advertised to contain 3BN faces.

Recap:

- Predictive models are statistical processes which "learn" how to predict results from training data.
- Images are stored as collections of pixels, which may be scanned and analyzed by repetitive simple processes.
- A convolutional neural network is a combination of image processing and a trained neural network model.
- CNNs for facial recognition may have high accuracy, but accuracy decreases when ideal conditions are not met.
- Services offered by vendors like Amazon and Microsoft include the recognition algorithms, but not the underlying database. That must be supplied by the user.

Thank You!