lecture one

on artificial intelligence and face recognition

xiaowan : sh-iao one

x.yi@arts.ac.uk

https://vimeo.com/user65401583

xiaowan-yi.com

(hopefully) Get you interested in AI

Clarify some notions around AI

Rethink about numbers

Have an awesome face detector app on your iphone made by yourself

“a boring lecture on Thursday afternoon and everyone is late”

- generated by https://huggingface.co/spaces/stabilityai/stable-diffusion

AI is fundamentally interesting…

Why?

How I get interested in AI

Cozy environment

Curious about ourselves (mind-blowing moments)

It’s NOT hard (easy to follow because it closely connects to our daily experience)

what is this sound?

https://www.youtube.com/watch?v=bbLDfueL7eU

bringing the vid…

how do you feel?

Once upon a time…

Pattern recognition: find regularity, enable prediction making 🔮

travel back to our era…

what do we have now?

The significance of pattern recognition to humanity

“

We distinguished predator from prey; and poisonous plants from nourishing ones - enhancing our chance to live and reproduce, and passing on our genes. We used pattern recognition in astronomy and astrology, where different cultures, recognizing the patterns of stars in the skies, projected different symbols and pictures for constellations. We used it to predict the passing of the seasons, including how every culture determined that the passage of a comet was taken as an omen.”

— “When Knowledge Conquered Fear”, third episode of the documentary tv series Cosmos: A Spacetime Odyssey

Pattern recognition as an essential part of our experience:

Guess the bonfire sound, Get on the right tube line(daily task solving), read my handwriting(language), appreciate music(art), etc.

Name more…

Artificial Intelligence

what is it??

“intelligence, made by human™”

it is still an unfinished goal

we usually hate artificial something

intelligence{

Intelligence is a big bag word

It includes the ability to solve complex problems or make decisions with outcomes benefiting the actor

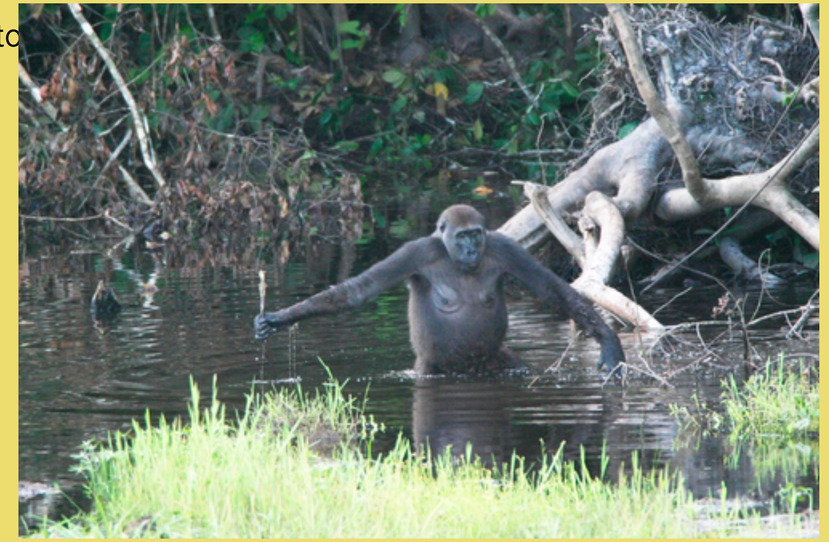
and many more…

Is intelligence exclusive to human?

We have:

Gorilla uses a stick to test depth of water

https://en.wikipedia.org/wiki/Tool_use_by_animals#:~:text=Chimpanzees%20are%20sophisticated%20to



Fish makes art:

https://www.youtube.com/watch?v=VQr8xDk_UaY

Dog talks:

https://www.youtube.com/watch?v=QKQK7EIcq9Y

~~What is intelligence?~~

Intelligence is always specific, most of time we are interested in human intelligence (for simplicity, we omit human in the rest of slides)

What are we able to do with intelligence?

Human task solving, art making , etc. what else?

If we manage to build the machine to be able to solve tasks and make art, can we say we have achieve AI?

Intelligence = task solving + art making?

emotion

Thinking about intelligence is a way of appreciation

Simple things are hard

“Be the subject of your own thoughts”

```
}
```

artificial intelligence{

Here are some human-made concepts of subjects:

Engineering science:

make tools🔭🚒✈️ — life gets easier 🫶

Natural science:

discover and explain phenomena🧪💯 — curiosity satisfied ✌️

These two are actually intertwined

So does AI!

AI is a tool (like engineering science)

- how to use tool (CoreML🍎)

- how to make tool (CreateML🍎)

AI is also our attempt to understand intelligence better (natural science)

“ We don’t think we have understood something unless we can build it fRoM sCrACh.”

From scratch: use machine which is created by human, not other organic beings we don’t fully understand yet

Then…

Where shall I start with if I want to make machines to be intelligent, aka to be able to solve tasks and make art?

Recall:

Pattern recognition as an essential part of human experience:

Get on the right tube line(daily task solving), read my handwriting(language), appreciate music(art), etc.

Let’s make the machine to do pattern recognition! 🔮

Let’s make the machine to do pattern recognition! 🔮

Do you know how cool it is?

Few hours of computations by our little metal box

~=

Hundreds of thousands of years of human evolution and experience accumulation

Shout out to Patrick Winston https://youtu.be/Unzc731iCUY?t=2323

Here are what we've talked so far:

✦Pattern recognition is everywhere and amazing

✦We want to understand intelligence by prototyping it using human-made machine ("AI")

✦One intermediate goal of prototyping intelligence is to make machine to do pattern recognition(because we guess that pattern recognition is an essential part of intelligence)

🧐How does this image-generating AI demonstrate its pattern recognition ability?

“a boring lecture on Thursday afternoon and everyone is late”

- generated by https://huggingface.co/spaces/stabilityai/stable-diffusion

```
}
```

Insert self introduction here :)

Research https://anonymous84654.github.io/RAVE_anonymous/

Drum and AI https://vimeo.com/93213203

Story time

What are AI researchers like?

“sleeping”

Geoffry Hinton, the godfather of DL, is recently taking inspiration from what the purpose of sleeping is and why we have dreams… 🔮 https://www.youtube.com/watch?v=2EDP4v-9TUA

“a lucid dream”

- generated by https://huggingface.co/spaces/stabilityai/stable-diffusion

Studying AI is a thought-provoking process

And it will get us to know ourselves better (lots of fun facts to come…)

Enjoy !

Noodling time..

🤨

Intelligence is not exclusive to human

Other species can also make and use tool, solve tasks and create art…

And now that machine can do something too

What make us us then?

“humanity ~= human capability - artificial intelligence"

🤨

Is it true that we assume machine intelligence is always a subset of human intelligence?

Is it possible if machine can actually do things that we intrinsically can not do? Like some of machine intelligence capabilities are beyond that of human intelligence ?

Representation{

What is representation?

“descriptor” “features” “characteristics”

Why do we need to have the notion of representation:

1. It is inevitable as a result of our “flaw”, more on this later

2. It is also a powerful tool towards task solving

What is apple?

Pattern recognition question 1

What is apple as a fruit vs. Apple as tech company?

(Efficient) representations:

- edible ?

- Upper case?

- etc.

Pattern recognition question 2

What is apple vs. pear?

(Efficient) representations:

- shape ?

- its taste ?

Good representation simplifies our task

To excel at pattern recognition ~= To find a good representation

ANIME time ! DOMAIN EXPANSION

https://www.youtube.com/watch?v=nmvkhLz8t7I

(Efficient) representations:

- shape ?

- its taste ?

Meet “papple ” …

https://www.theguardian.com/lifeandstyle/wordofmouth/2012/may/21/the-papple-tasted-and-tested

To get out of ambiguity, just ask about the context

Why do you want to know if it is an apple or pear?

Representation is contextual

Depends on the problem given, different tasks have different efficient representations

Perhaps we can never describe/represent one thing as it is with nothing less nothing more…

our natural language doom to fail (our “flaw”)

Representation

- descriptive, captures some characteristics
- contextual, task-dependent
- perspective, always partial

Another related notion

Abstraction:

Taking away irrelevant details, reducing the representation to essential characteristics

Joel’s slides on “what is computational thinking” https://jgl.github.io/DiplomaInAppleDevelopment-AutumnWinter2022/codingOne/lecture_01.html#42

Lots of things are connected. Studying AI is a brilliant manifestation of computational thinking.

Numbers{

What is the domain where we can have almost perfect representation (aka without ambiguity)?

I have three pens

what is “I” what is “have” what is “pen” 🤯

what is “three” 🤗

- Numbers
  - Count
  - Measure
  - ..?

007

- Numbers
  - Count
  - Measure
  - Label

We always need a “protocol”(like an agreement on how to interpret numbers, or like a dictionary for looking up number’s meaning) when using numbers in real life.

Though numbers provide an almost perfect representation domain, it is “fictional”🥲

In real life, we don’t see numbers on their own walking on the street

When we encounter numbers in real world, there are always real-world meanings attached to numbers

We always need an interpretation guide(“protocol”) when using numbers in real life.

Why do I want to talk about numbers?

Our human-made poor machine can only deal with numbers 🫠

Numbers can introduce maths, which is our DOMAIN EXPANSION 💥

It is SUPER important to grasp the idea of using (numbers + protocol) to represent things, for doing fancy AI stuff😈

}

End of noodling,

Starting ordinary lecture mode…

- promise 🎉

by the end of this unit, you will:

- know how artificial intelligence works in practice
- make a wide range of ios apps including face detection, speech recognition, activity classification, etc…
- have more thoughts on artificial Intelligence as a cultural concept

scope of this module

- describe how machine learning works in practice (Knowledge)
- construct applications with the Core ML framework (Process)
- discuss artificial Intelligence as a cultural concept (Enquiry)

for each lecture, we will go

describe☕ -> construct🔧 -> discuss🍵

lecture plan today

✦machine learning model introduction☕ 40 mins

✦face detection introduction☕ 40 mins

✦write your awesome face detection ios app🔧 1 hr

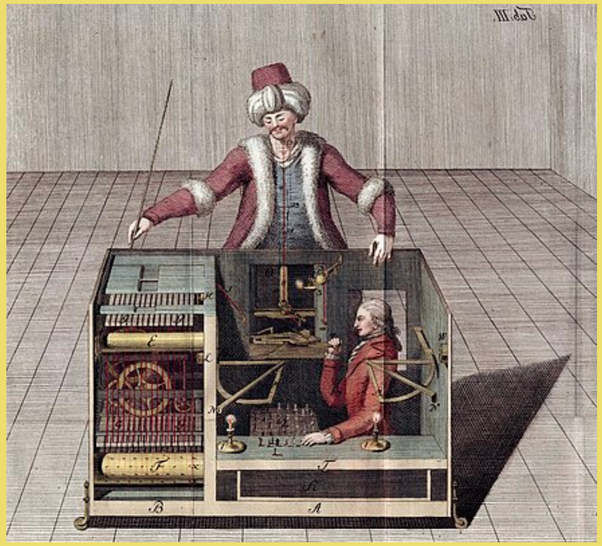
✦discussion🍵

and breaks in-between…

- AI? ML? 🧐
- interchangeable throughout this unit

- Machine “learning"

- Mechanical Turk
  - https://en.wikipedia.org/wiki/Mechanical_Turk



- “Learning”: we train machines to solve tasks, machines are not quite autonomous

- AI is a general term (cultural impact, etc.)
- ML is a technical term (algorithm, code framework, etc.)
- Deep Learning…? Neural Networks…? 🥺
- DL(NN) ⊂ ML ⊂ AI

“interchangeable”

before diving into how machine learns…

how do we learn??🧩

- ✦ questions

  how you can reach out to the world and gather knowledge

- ✦ intuitions

  how you can internalise the knowledge efficiently

how do we learn? - questions

- whatever questions you have in mind — interrupt me anytime !
- having a question means you have already known enough to know what is not known[1]
- i’m also learning from your questions

how do we learn? - intuitions

- intuitions connect you from academic jargons to daily real life
- some intuitions on ML can be gained just by introspection
  - i’m also here to help by sharing mine
- a lot of AI developments are largely inspired by what we think of how ourselves are being put together

  “attention mechanism”

machine learning model

before diving into machine learning model…

“information era” 

information we receive from the world are mainly from four categories:

- image (video)
- text (language)
- sound (music, speech)
- numbers(the weather in degree celsius , your birthday, etc.)

can you think of any information that is not from the four categories?

there are…

Terminology used by AI nerds

data category = data **modality**

my mind-blowing moment:

information from any of these three categories (image, text and sound) can be represented by just a bunch of numbers

using numbers only🤪

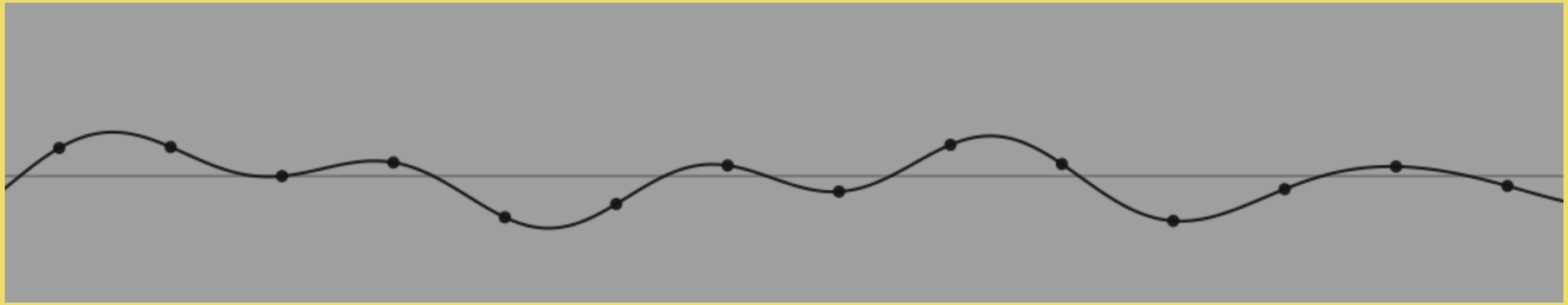
image in numbers:

- two numbers for its width and height (how many pixels)
- for each pixel, what the rgb values are

language in numbers:

- we will talk about this later
- but for now just think about when you looking up a word in a dictionary📕 using page number and index
- (also math itself is a language….)

sound in numbers:



this is a wav file of a drum beat, screenshot with a lot of zooming in

each dot represents a number

why do we care about represent things in numbers?

- because **computers** can only deal with numbers

- because with numbers we can do our DOMAIN EXPANSION aka math

machine learning model

? ? ?

what is a model ?

✦some common ML models:

- ✦face detection model 😤
- ✦dog-or-cat image classification model 🐶😼
- ✦speech recognition model 🗣️💬
- ✦language translation model 💬💬

What do they have in common?

they are all nice tools but every one sounds very different from each other!

- Though sounding different, they share the same structure of what a model is (recall a “class” or a “protocol” in swift)
- each ML model takes in **input**, does some process and generates **output**
- “doing some process” is usually where the elegant maths, computations and perhaps confusion happen
- **However, input** and **output** are the important specifications of a ML model

“what are the input and output of this model” is always the **first** question to ask when you try to understand a model ❣️

it also helps answering this question: what does this ML model do

- what are the possible input and output?
- - data, aka information
- ML models take in information, do some process, and generate (hopefully useful) information
- remember those four categories of information?
- our candidates of input and output are:
- image🩻, text📗, sound📣, numbers💯

“what are the input and output of this model” is always the **first** question to ask..

It also helps answering this question: what does this ML model do

try this…

what does a speech recognition model 🗣️💬 do?

try answer using “given a <your educated guess on the input>, the speech recognition model generates <your educated guess on the output>”

image🩻, text📗, audio🔊, numbers💯

and try this…

what does a dog-or-cat image classification model 🐶🐱 do?

try answer using “given a <your guess on the input>, the dog-or-cat image classification model generates <your guess on the output>”

image🩻, text📗, sound🔊, numbers💯

How to shepherd the meaning of a bunch of numbers in the output?

How do we know what each output number represent?

During training (next unit),

specify which output number means what (the “protocol”), this protocol will stay consistent across the life span,

the protocol of how to interpret each output number should be passed to model users

An example

Task context:

Use numbers to represent whether the image is a dog or cat

The number representation I come up with:

[0, 1]

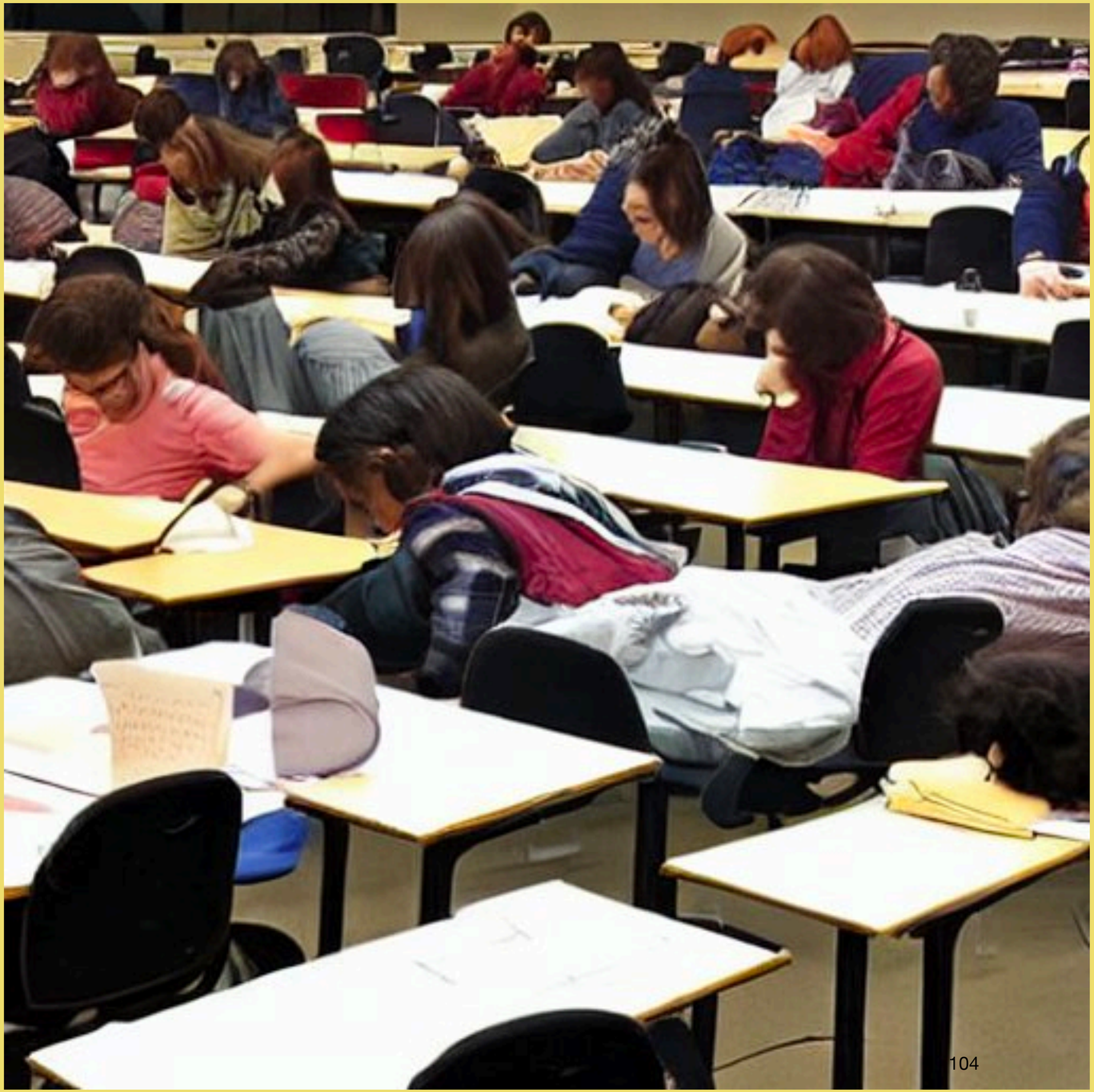
The protocol I’m going to pass around:

hello this is a protocol created by covfefe for the dog-or-cat numeric representation and I can bullshit whatever I want here as long as I explain how to interpret [0, 1] somewhere are you with me

The first number (with index 0) in this array represents the probability of this image being a dog image

The second number (with index 1) in the array represents the probability of this image being a cat image

now try this



output

input

“a boring lecture where everyone is sleeping”

- generated by https://huggingface.co/spaces/stabilityai/stable-diffusion

think of ML models as tools ⚒️

- how to make tools
  - we have not talked about how to make ML models, aka “doing the process” part
  - we will be looking at how to make ML models at a later time

✦how to use tools (our focus of this unit)

  - we need to know each tool’s specifications
  - what are the specifications of ML models?
  - - input and output ✨⚡🧨

coincidentally, apple ML framework has a similar division too

how to make tools - CreateML

how to use tools - CoreML

also coincidentally, the “input and output” thinking of an ML model manifests in how apple defines a ML model in its CoreML framework:

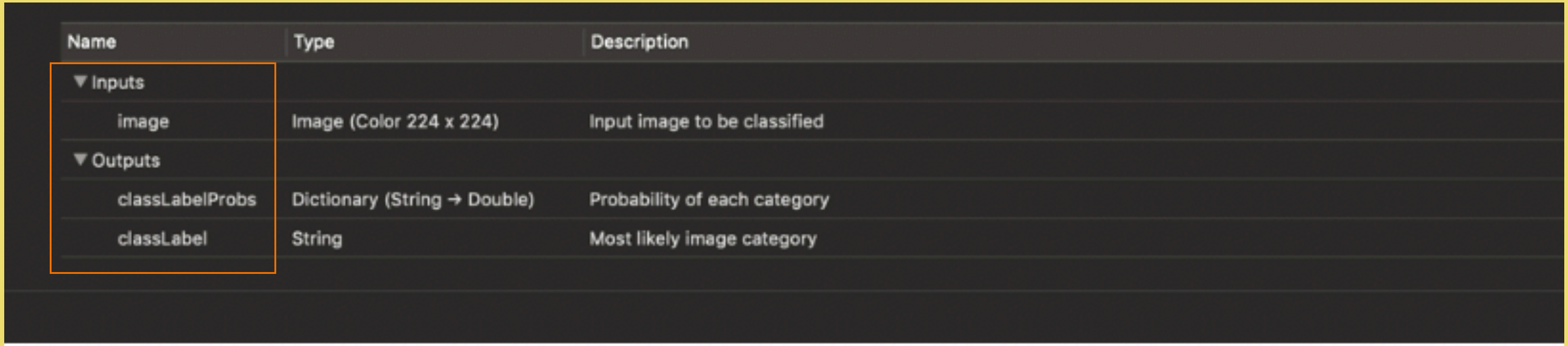


Figure 2-4. The MobileNet .mlmodel inside Xcode

can you find this “input, process and output” mechanism in us as human beings?

https://www.youtube.com/watch?v=X5fD0Evny4w&t=36s

end of machine learning model introduction question?

face detection model

while your memory is fresh..

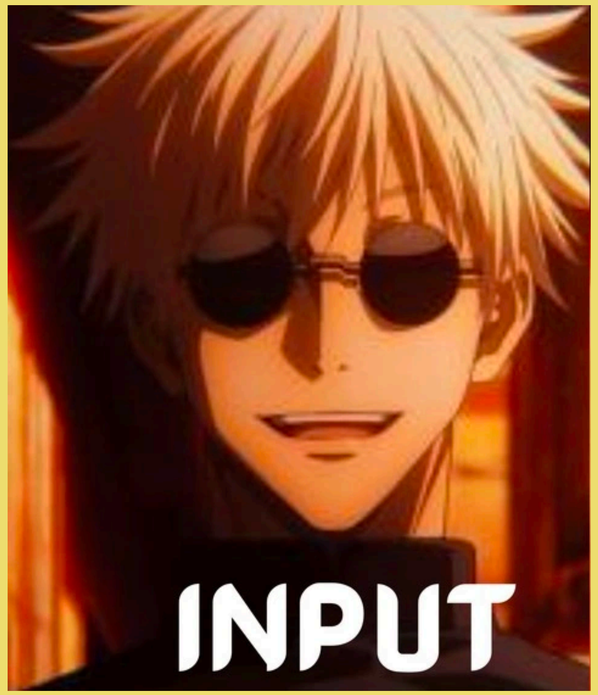
what does a face detection model do?

try answer using “given a <your guess on the input>, the face detection model generates <your guess on the output>”

image🩻, text📗, sound📣, numbers💯

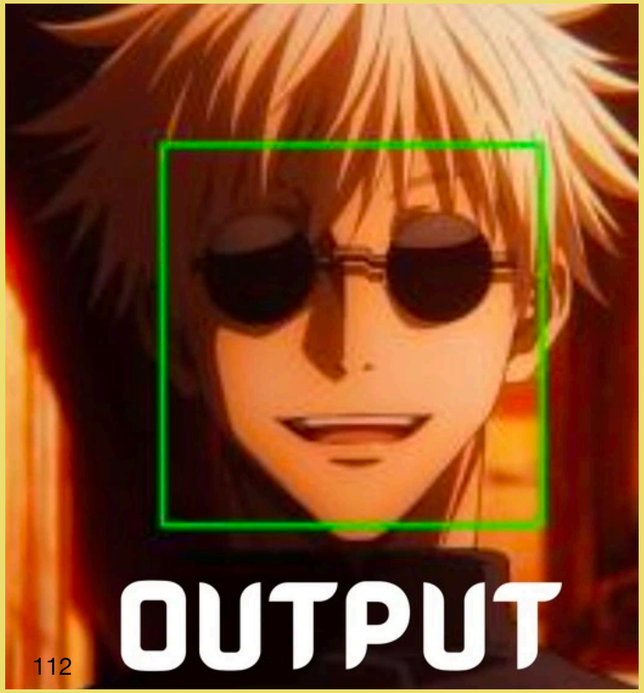
face detection model

given an image (with or without faces, could be any),



the face detection model generates

the detected locations of faces



what can we use the model output (detected face locations) for?

what can we use the model output (detected face locations) for?

counting how many faces are there ☝️

draw the detected face location bounding box on the image ✍️🔲

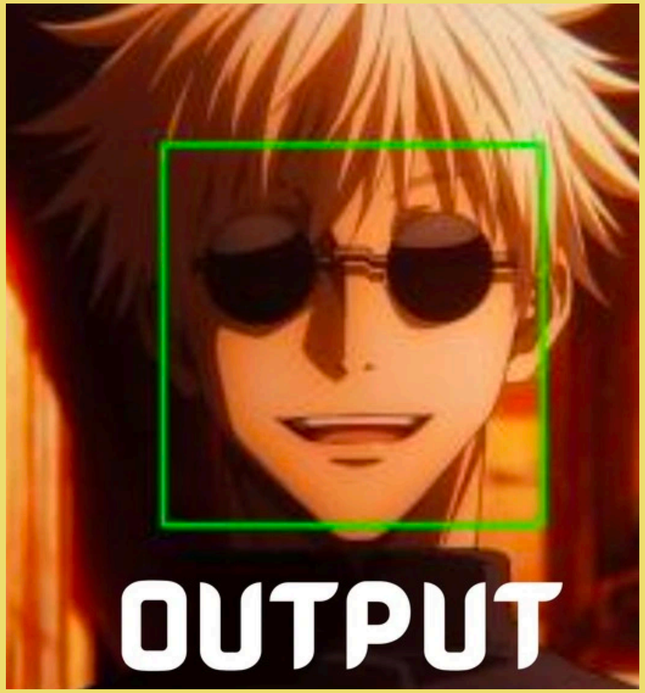
applying an emoji to cover the face🤪

we will be building an app to achieve all of these in a minute!!!

go to app construction now…

or if time allows we can dive a bit deeper into the face detection model introduction

❗ a face detection model does not generate output in the form of this nice green rectangular bounding box as you see



🤏 its output is actually a numeric representation of this green box (recall we can represent all those amazing stuff in numbers?)

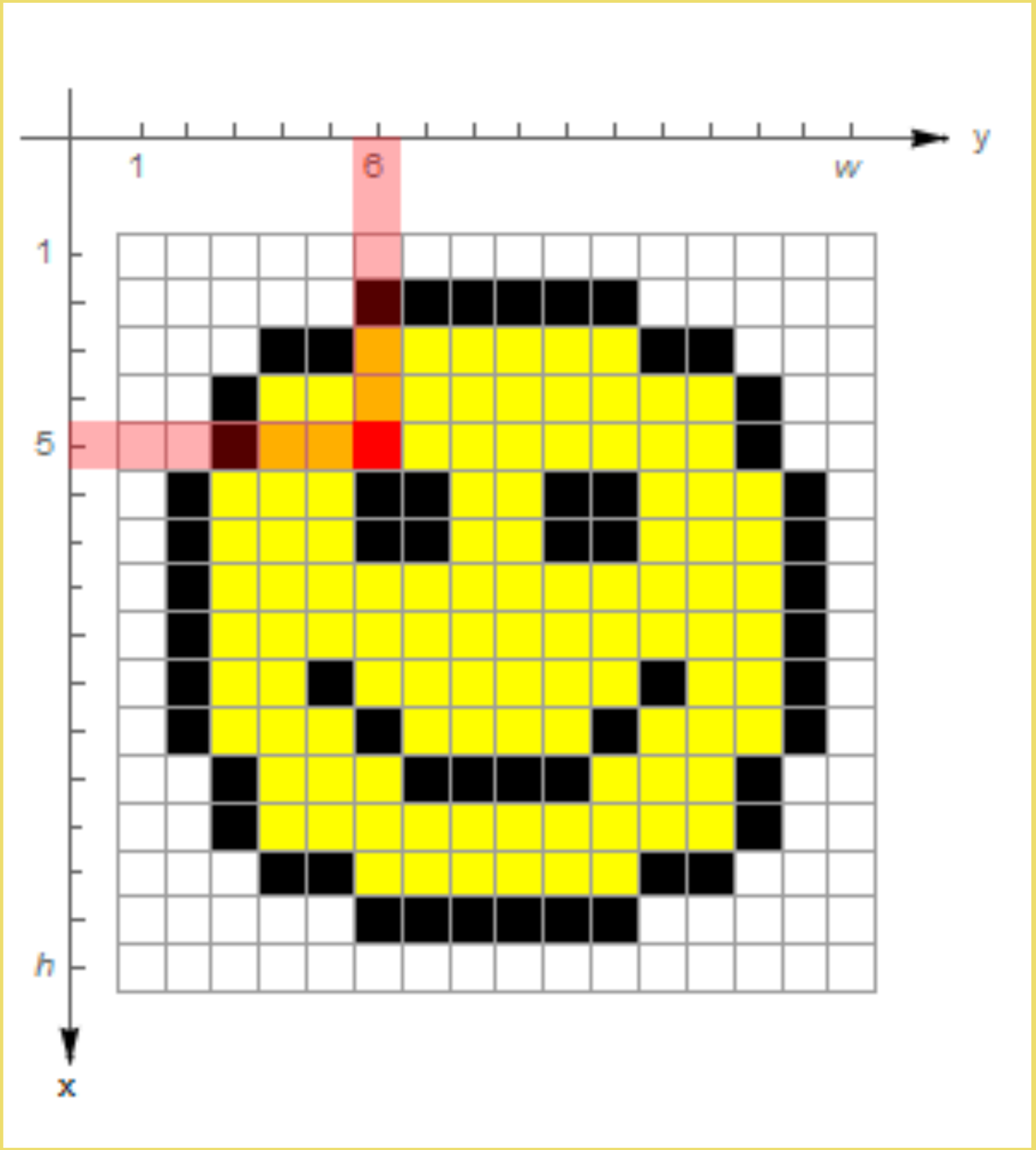
🥰 based on the number representation of the bounding box, we programme the computer to help us draw out this box

how is the detected face location, aka bounding box, represented in number?

an easier question

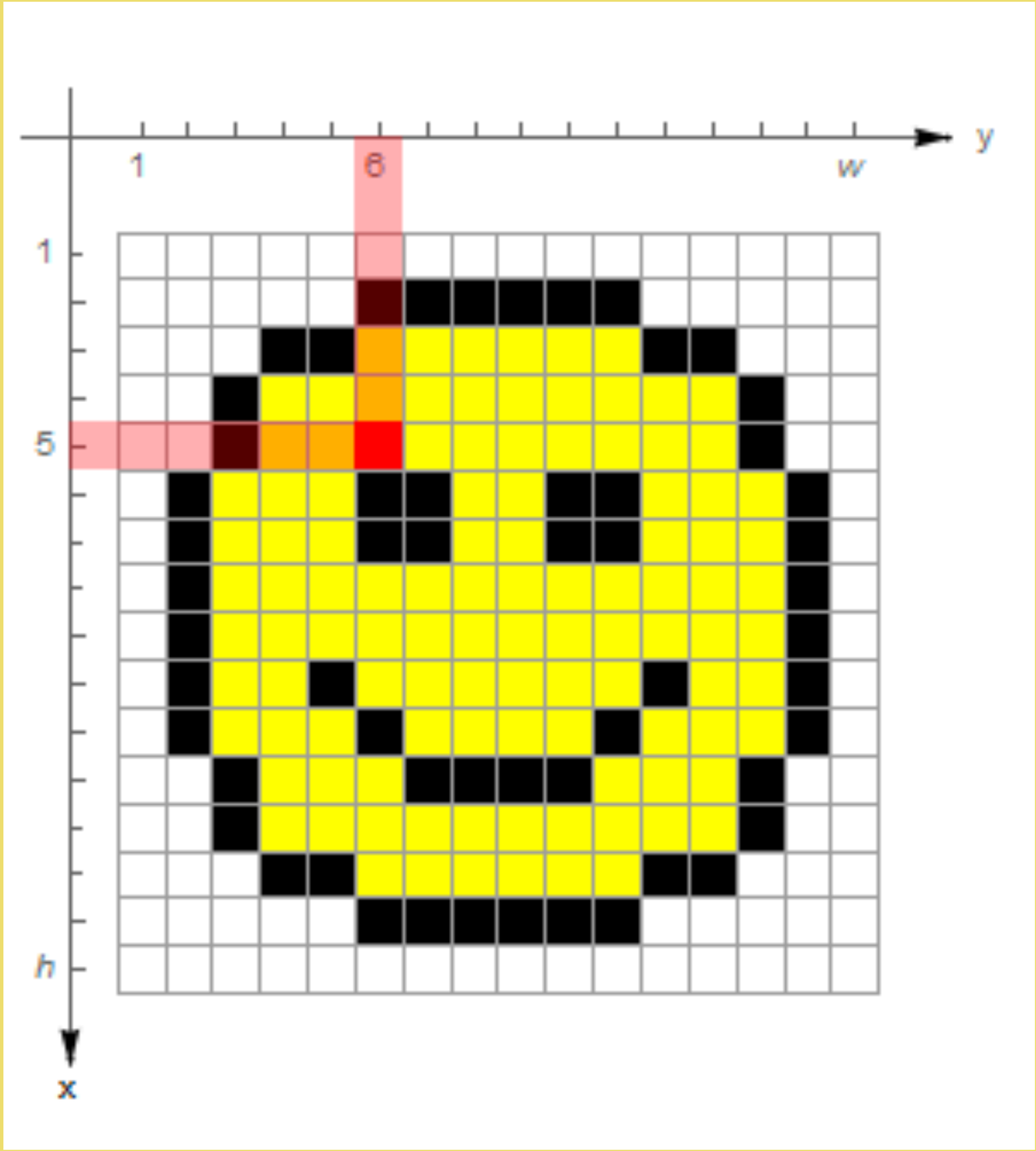
how is the location of a single point in an image represented in numbers?

- the location of a point is denoted by two numbers, aka coordinate (x, y)



- x represents the distance (in number of pixels) between the point and the left most edge

- y represents the distance (in number of pixels) between the point to the upper most edge



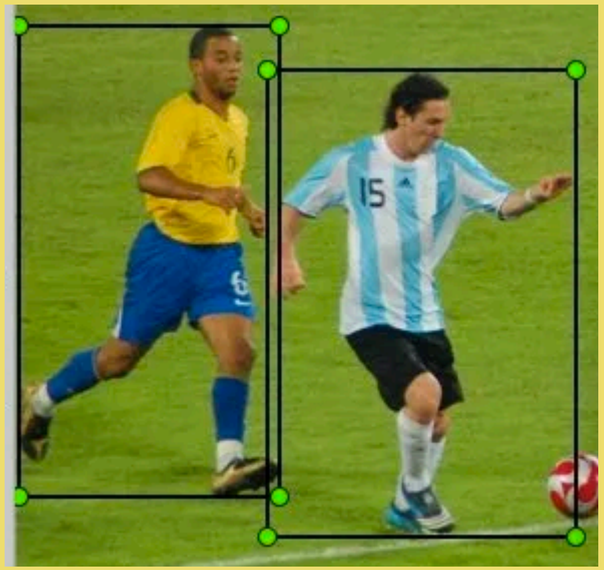
now that we know how a single point is represented in number ☑️

a bounding box is nothing but a combination

of its four corner points

once the four corners is known, we just sit and let the computer to draw the lines for us ☕

Example on how to represent one bounding box in numbers:



Location of upper-left corner: [0, 0]

Location of upper-right corner: [20, 0]

Location of upper-left corner: [0, 40]

Location of upper-left corner: [20, 40]

One bounding box: [[0, 0], [20, 0], [0, 40], [20, 40]]

Don’t forget the protocol:

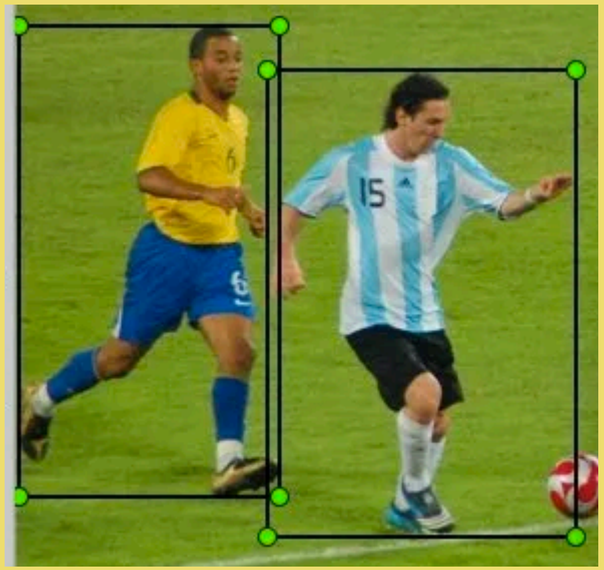
<insert your educated guess here>

Noodling time:

Given the representation of one bounding box: [[0, 0], [20, 0], [0, 40], [20, 40]]

(with <protocol same as in last slide>)

Can we infer the width and height of this box?



Noodling time:

do we really need all four corners’ (x, y) coordinates to be able to draw the bounding box? 👀

some face detection model can do more than just figuring out where the outline of face is…

recall when you see someone’s face image from a book and you tend to look into their eyes for one second?😳

some face detection models can do something similar…

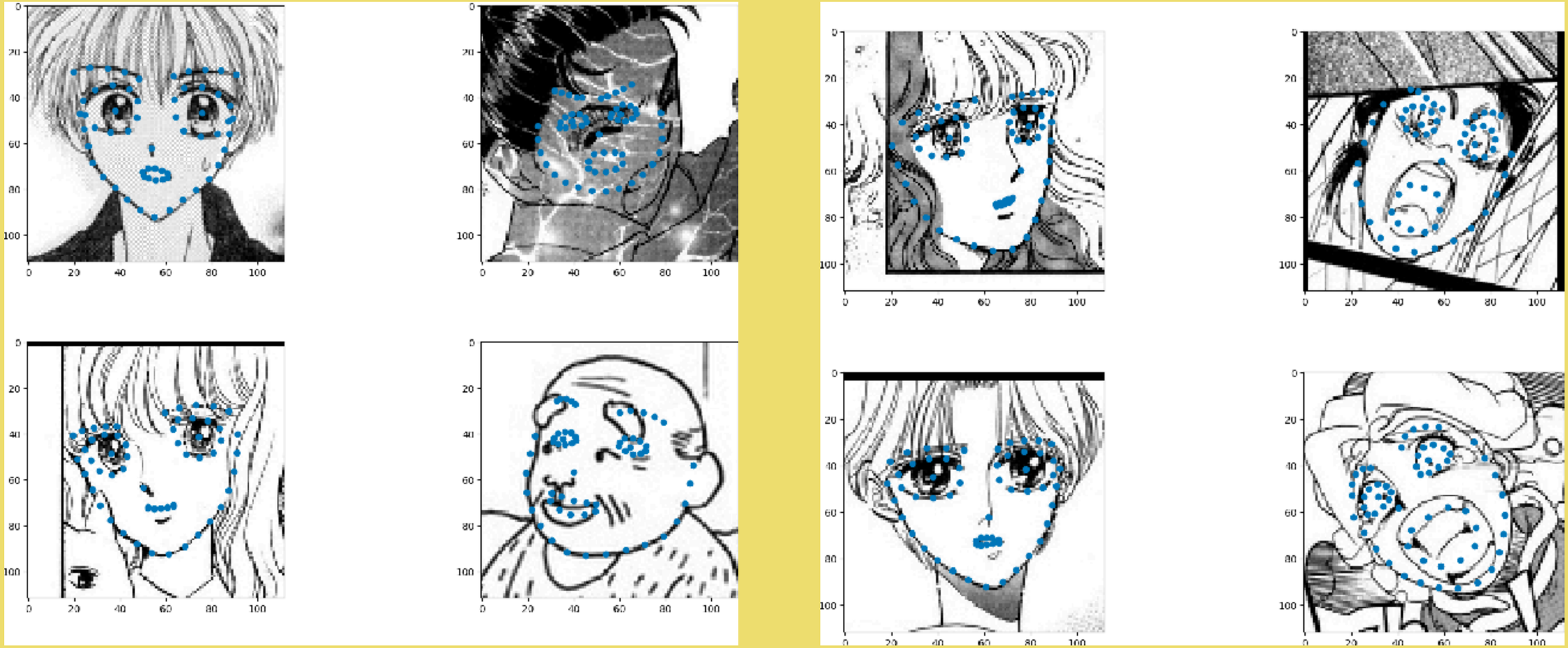
they can find the exact locations of eyes

and nose-tips, and many more…

these points are called “landmarks”

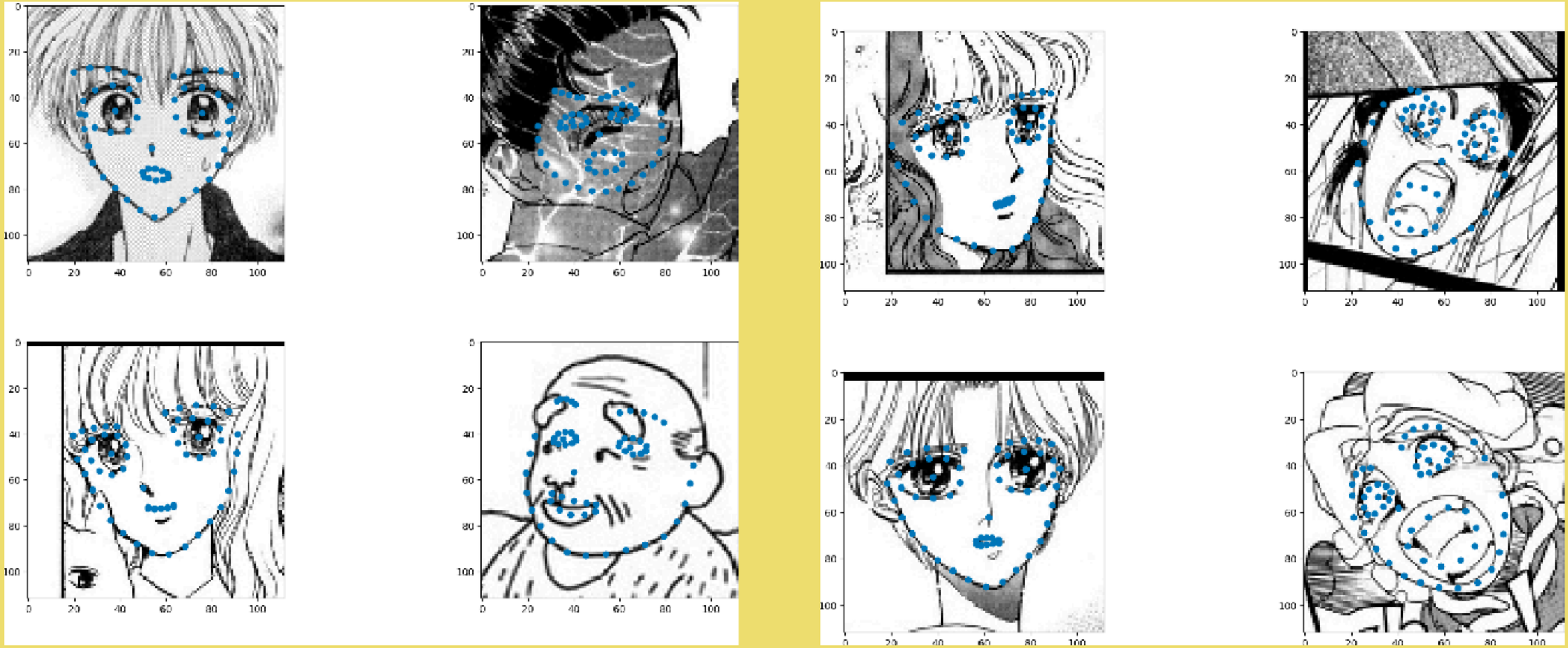
check what landmarks apple’s model can find https://developer.apple.com/documentation/vision/vnfacelandmarks2d

here is a face detection landmarks output of manga images [2]



each landmark is represented by its coordinates

a set of landmarks means a set of coordinates



Example on how to numerically represent landmarks numbers:

Location of right eye mid point : [30, 20]

Location of left eye mid point: [50, 20]

Location of nose tip point : [40, 40]

Location of upper-left corner: [20, 40]

and other points of interest…

Landmarks: [[30, 20], [50, 20], [40, 40] … etc.]

Don’t forget the protocol:

Arrays are in the order of right eye mid point, left eye mid point, nose tip point, etc.

what do we need the landmarks for?

the bounding box can only tell if some region has a face, regardless of its rotation🙃

landmarks can tell us the rotation of the face

we need this information to perfectly overlay emoji

finding faces may seem trivial for our visual system,

it used to be a hard task for machines

we can locate face and landmarks in one go within a blink, for machines finding landmarks is another level of difficulty to achieve

“simple things are hard”

thankfully when coding an iOS app, we just need to type in the right function that’s all

detecting bounding boxes:

VNDetectFaceRectanglesRequest()

detecting landmarks:

VNDetectFaceLandmarksRequest()

By calling VNDetectFaceRectanglesRequest() or VNDetectFaceLandmarksRequest()

We are retrieving the output of apple’s awesome face detection model

Question for later:

Where do we feed input to the model?

end of face detection model introduction question?

construction time 🔧 !!!

https://github.com/XiaowanYi/MLOne-DiplomaInAppleDevAW22-Lec-01

preparation 1: which Xcode version are you using?

preparation 2:

there will be mostly cutting and pasting from the textbook

don’t be scared — you don’t have to comprehend every single line

preparation 3:

to get the cutting and pasting right, pay attention to which function or which object you are pasting into

“the scope”

let’s start the project by open Xcode:

Create a new Xcode project

iOS -> App -> Next

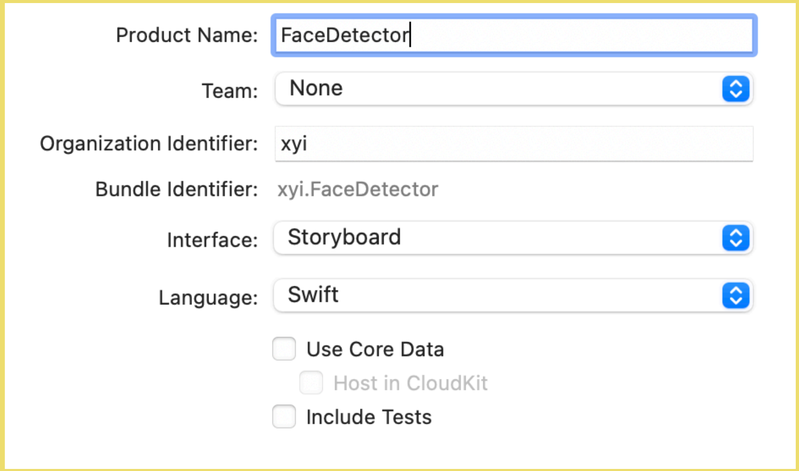
step a:

❗ select SwiftUI from interface dropdown menu

Use core data and include tests unchecked (not important for this project)

Click next and select a folder:

Good practice: creating folder in a designated working folder



step b:

❗ magic dust 1:

Info -> custom iOS target properties
-> “+” on any row -> select “privacy
- camera usage description”

step c:

let’s look at textbook P171- 173 step 3 & 4

don’t worry about errors notifications, they will be resolved as we progress

have you seen the familiar VNDetectFaceRectanglesRequest() ?

step d:

let’s move to textbook P174 step 1, 2 & 3

note we are moving to a new file (Views.swift)

this is defining ui buttons we will use later

step e:

let’s look at textbook P176-177 step 4, 5 &6

❗ correction 1:

in step 4 first line struct Main View

It should be struct MainView (remove the space in-between)

❗ correction 2:

in step 4 second line private let image: UIlmage

It should be private let image: UIImage (both are capital I not lowercase)

step f:

let’s look at textbook P177-181 step 7

this is a new and long struct, be careful

step g:

let’s look at textbook P181-182 step 8

this is for handling rotations

recall: when do we need rotations? 🙃

step h:

moving to file ContentView.swift file

step i:

let’s look at textbook P182-184 step 9, 10, 11

ui stuff

In step 11 the line with:

.navigationBarTitle(Text(“FDDemo"),

you can change the text string to be your own app name

step j:

let’s look at textbook P184-185 step 12, 13

step k:

let’s look at textbook P185-187 step 14, 15, 16

note from step 14 we go out of the scope of extension ContentView{} and pasting codes directly

step L:

let’s look at textbook P187-188 step 17, 18, 19

note from step 14 we go out of the scope of extension ContentView{} and pasting codes directly

step m:

❗ patch 1:

continue on step 19 , add the following function into struct ContentView: View {}

```
private func controlReturned(image: UIImage?) {
        print("Image return \(image == nil ? "failure" : "success")...")
        self.image = image?.fixOrientation()
        self.faces = nil
    }
```

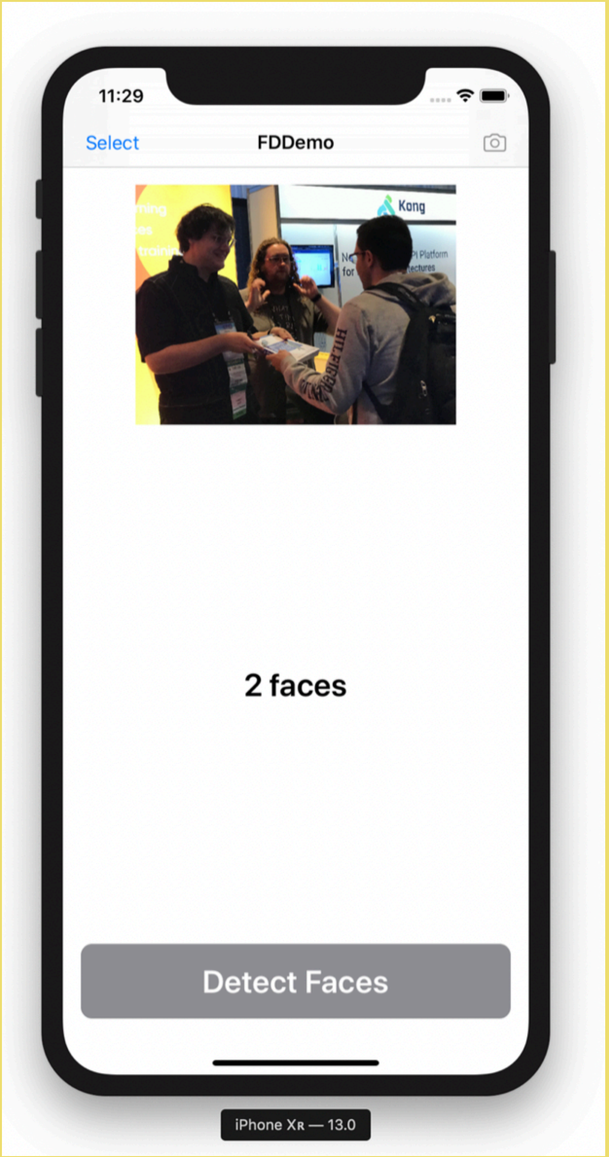
step n:

❗ magic dust 2:

add placeholder and icon to your Assets

(drag and drop to Assets in navigator)

building time !!!



next: draw the bounding box

step o:

let’s look at textbook P190-192 step 1, 2, 3

we are moving to file Faces.swift

here the codes corresponding to draw out the bounding box

you can customise the box colour in step 3
context.setStrokeColor(UIColor.red.cgColor)

step p:

let’s look at textbook P192 step 5

we are updating the entire getFaces() to incorporate the box drawing function (drawOn())

building time !!!

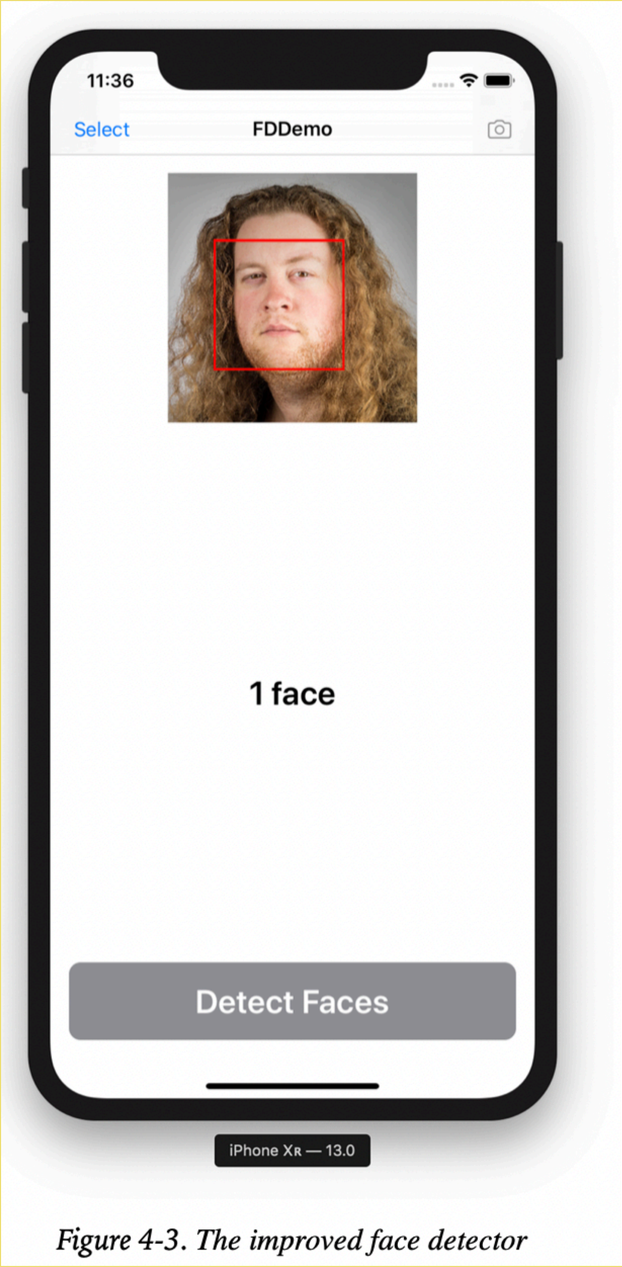


Figure 4-3. The improved face detector

next: applying emoji on top

we’ll only be working on the file Faces.swift

step q:

let’s look at textbook P197-199 step 1

for image rotation

step r:

let’s look at textbook P199-203 step2

recall: landmarks

VNFaceLandmarks2D here represents all of the landmarks that Apple’s Vision framework can detect in a face.

step s:

let’s look at textbook P203-207 step3, 4, 5, 6

adding extensions on the global scope

step t:

second last step!

let’s look at textbook P207-210 step7

it is replacing the entire extension on Collection

basically it replaces the box drawing function with the emoji placing function

if no emoji is shown your editor, you need to copy paste the list of emojis from line149 - 157 here

https://github.com/AIwithSwift/PracticalAIwithSwift1stEd-Code/blob/master/Chapter%204%20-%20Vision/Face%20Detection/FDDemo-Improved/FDDemo/Faces.swift

step u:

finally!!!

in Faces.swift -> extension UIImage {} -> roughly 4th line

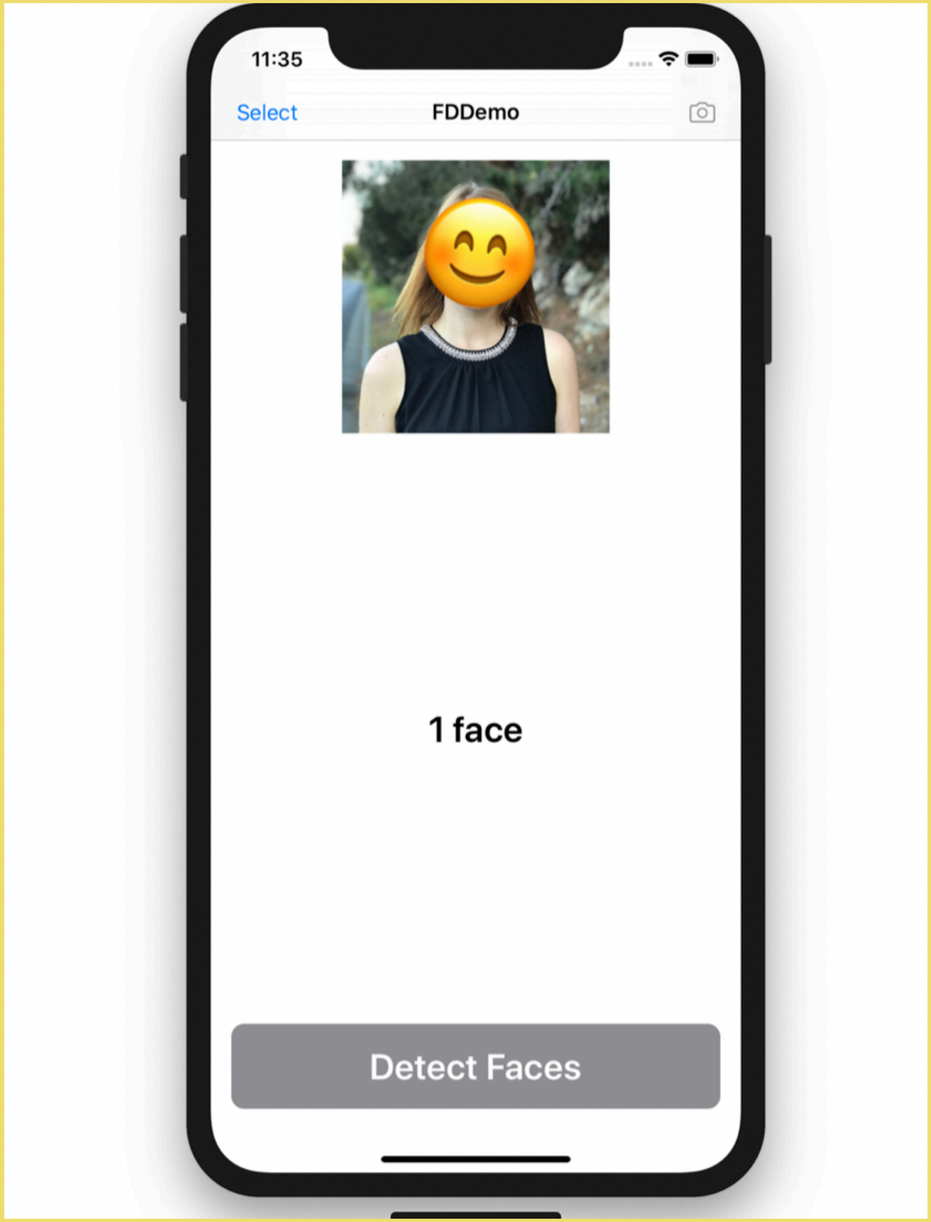
change let request = VNDetectFaceRectanglesRequest()

to

let request = VNDetectFaceLandmarksRequest()

(in order to switch from bounding box detection to landmarks detection)

building time !!!



congrats 🎉

don’t be scared about the code,

this lecture is about understanding the practical side of ML

as long as you get the idea of “using ML model output by calling the right function“

a gentle summary 😈

✦numeric representation 1 — image, text and sound can be represented using numbers with protocol☕

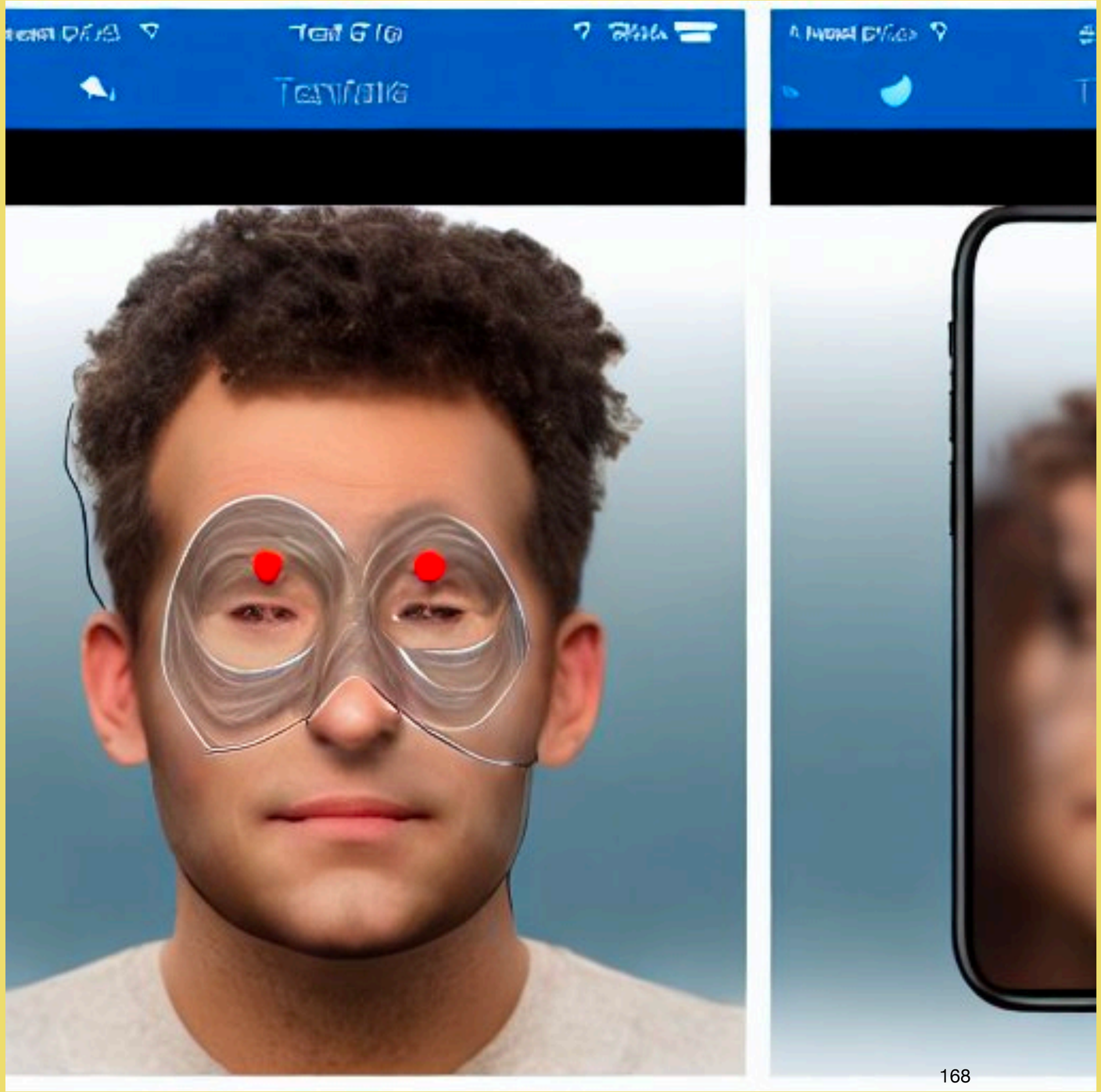
✦numeric representation 2 — a point location within an image is represented using coordinates(x, y)

✦numeric representation 3: — a bounding box within an image can be represented using coordinates(x, y) of its four corner points

✦input and output characterise a ML model☕

✦apple’s face detection model can output detected face bounding boxes through function VNDetectFaceRectanglesRequest() ☕

✦it can also output landmarks through another function VNDetectFaceLandmarksRequest() ☕



“a screenshot of an ios face detector app”

- generated by https://huggingface.co/spaces/stabilityai/stable-diffusion

References

ref 1: https://www.sciencedirect.com/science/article/abs/pii/S0022537179902007

ref 2: https://www.semanticscholar.org/paper/Facial-Landmark-Detection-for-Manga-Images-Stricker-Augereau/64cac22210861d4e9afb00b781da90cf99f9d19c

image ref https://animevyuh.org/face-detection-using-opencv/

image ref https://support.wolfram.com/25330?src=mathematica