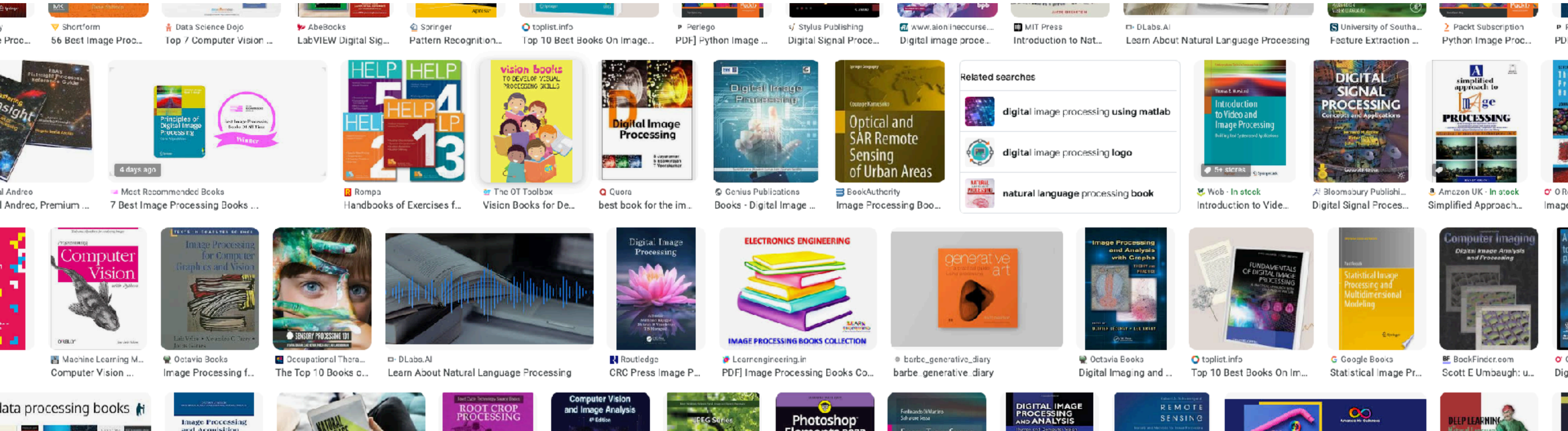
Adventures In Image Processing

Finding Patterns in Pixels

Chris Greening - iOSDevUK 2023

Maybe I should have picked something easier…



Don’t worry - it’s coffee time soon!

A Regular 2D Grid of Pixels

What is a Pixel (part 1)?

Pixel

Pix el

Picture element

Pixel

Picture Element

- “**pix**” - derived from “**pics**” which comes from **Pic**tures
  - First use of the word “**pix**” 1932
- “**el**” - **el**ement
- “**Pixel**” starts to be used to describe **pic**ture **el**ements around 1963
- Occasionally you may see “**Pel(s)**” used (**P**icture **El**ement)

An Image

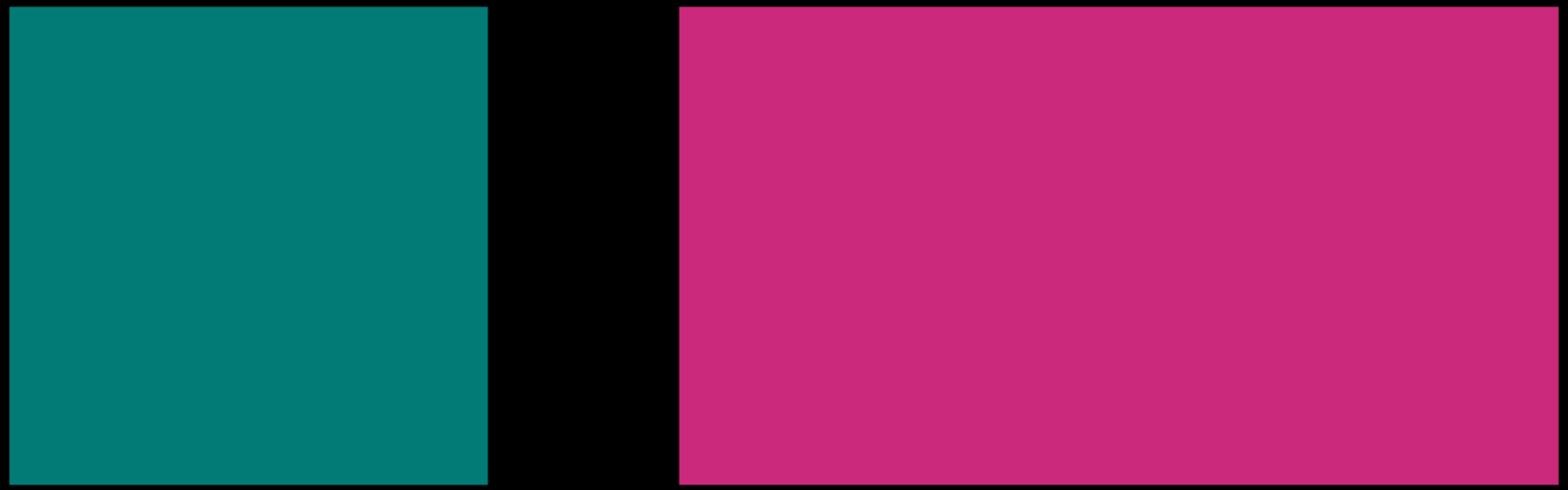
A Regular 2D Grid of Pixels

In the olden days

This was entertainment…

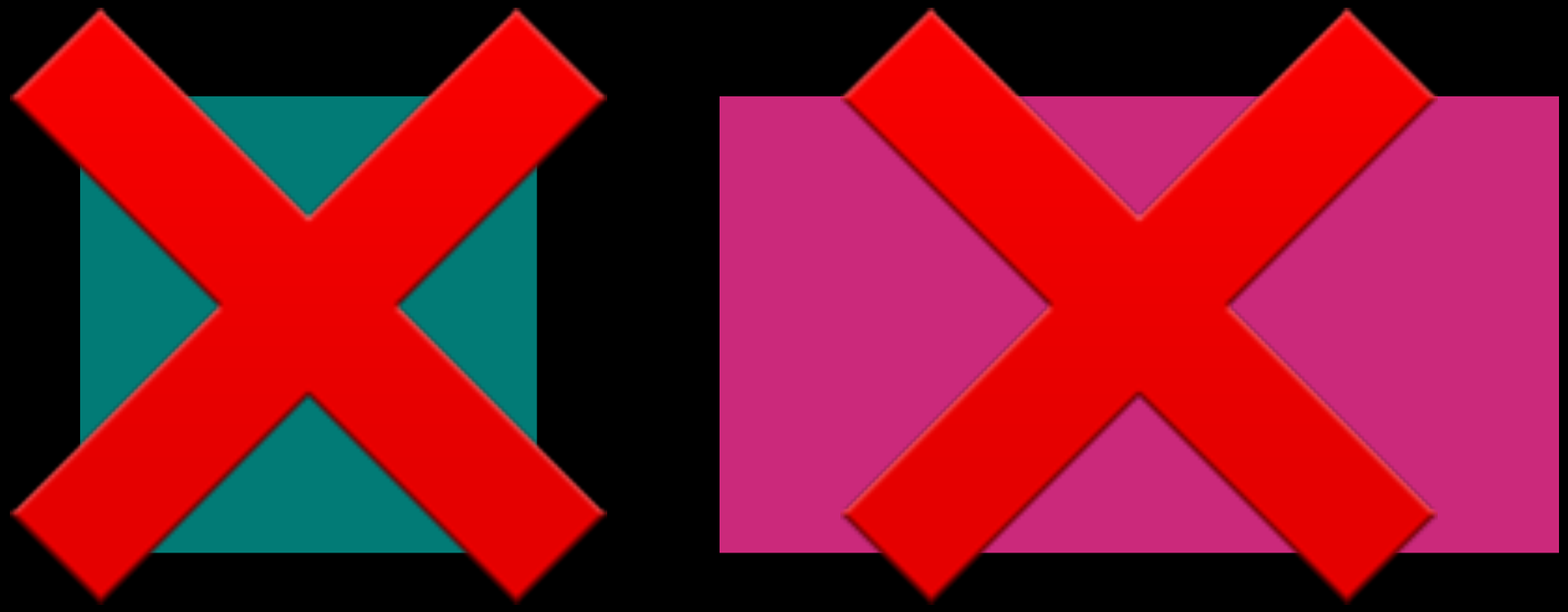
A Pixel is a Square?

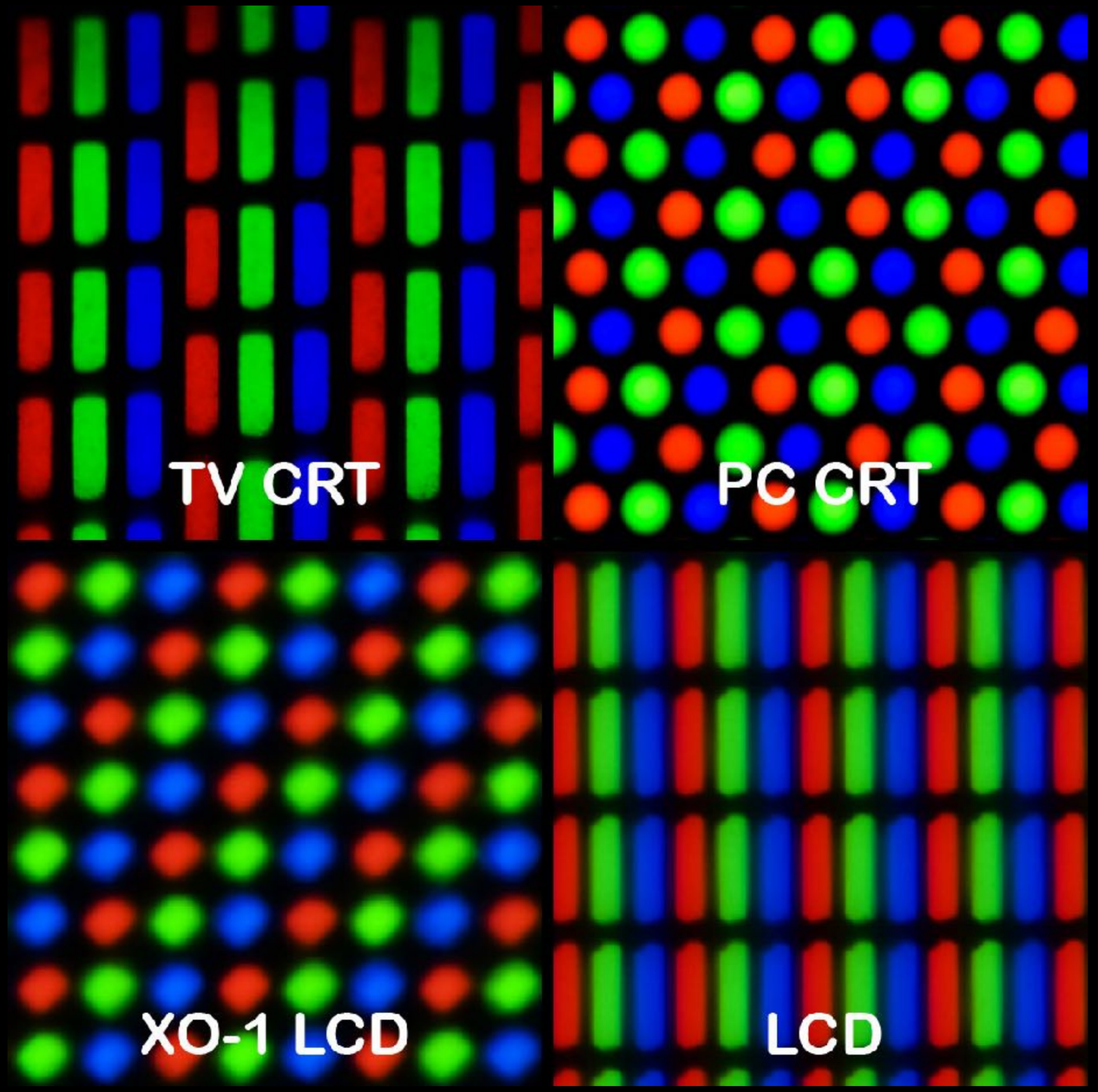
At worst it’s a rectangle…



A Pixel Is Not a Square

It’s not a rectangle either





By Pengo - Own work, CC BY-SA 3.0, https://commons.wikimedia.org/w/index.php?curid=3430548

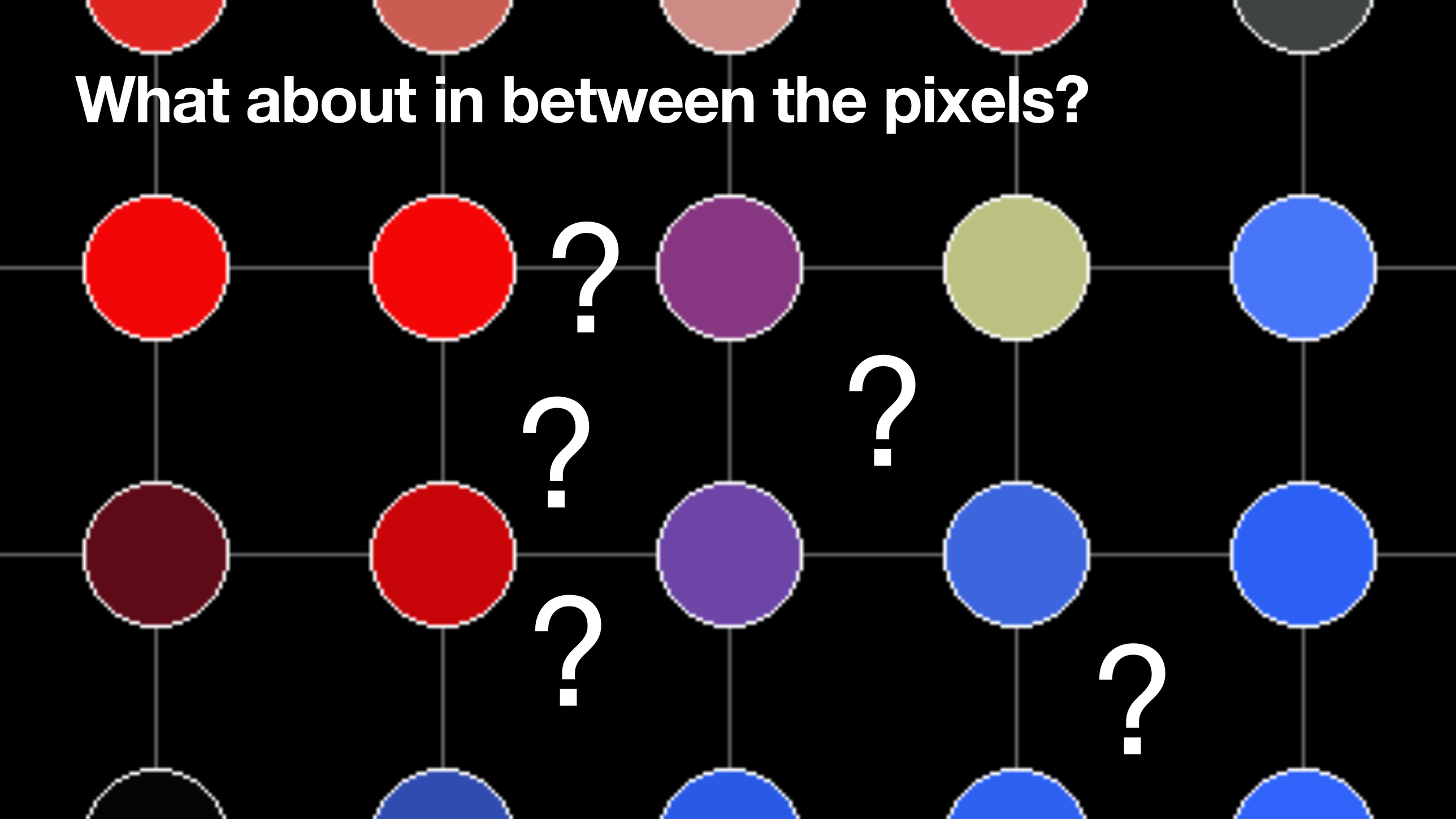
And don’t get me started on how images are captured…

So what is a Pixel?

Point Samples

What about in between the pixels?

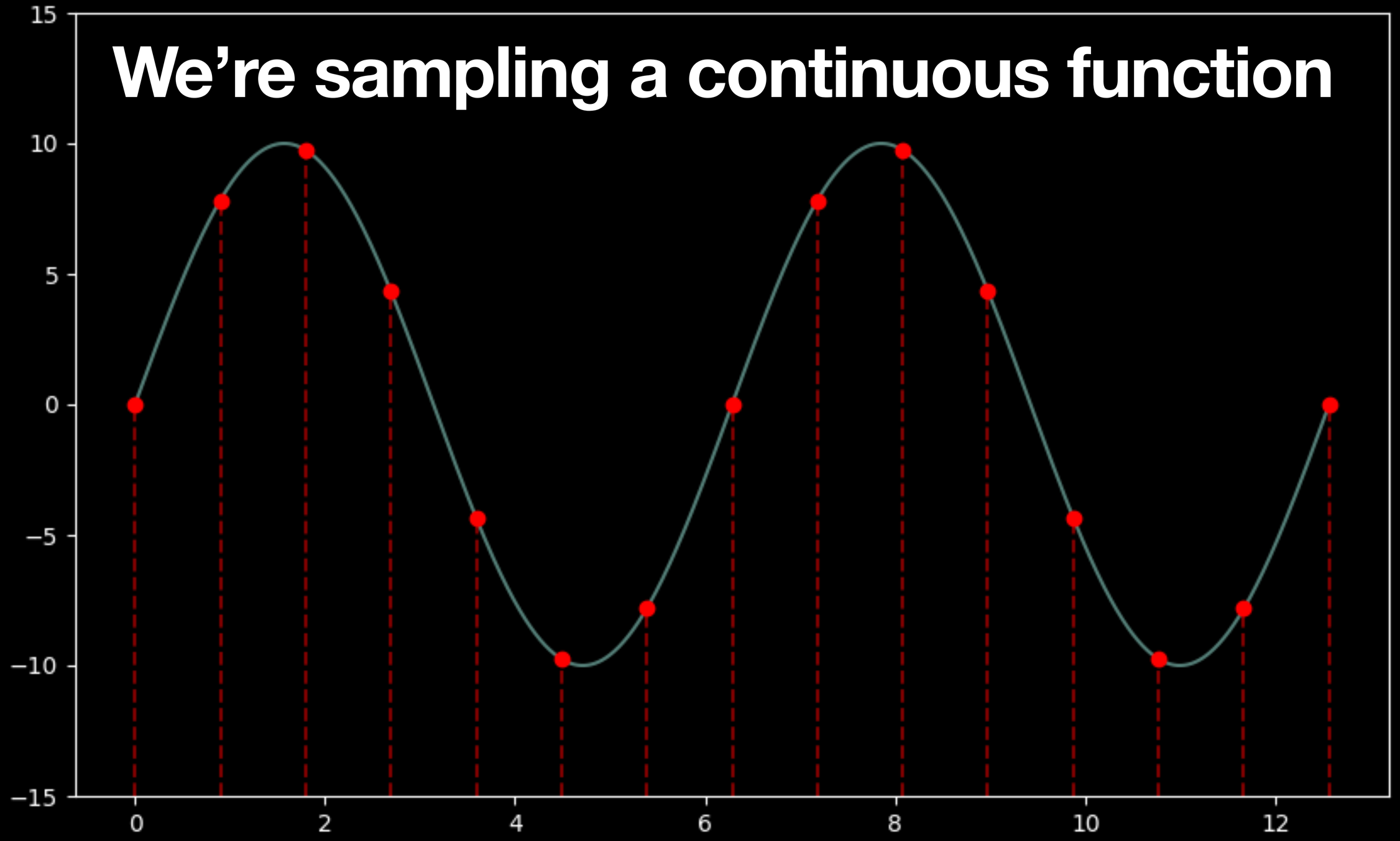
What about in between the pixels?



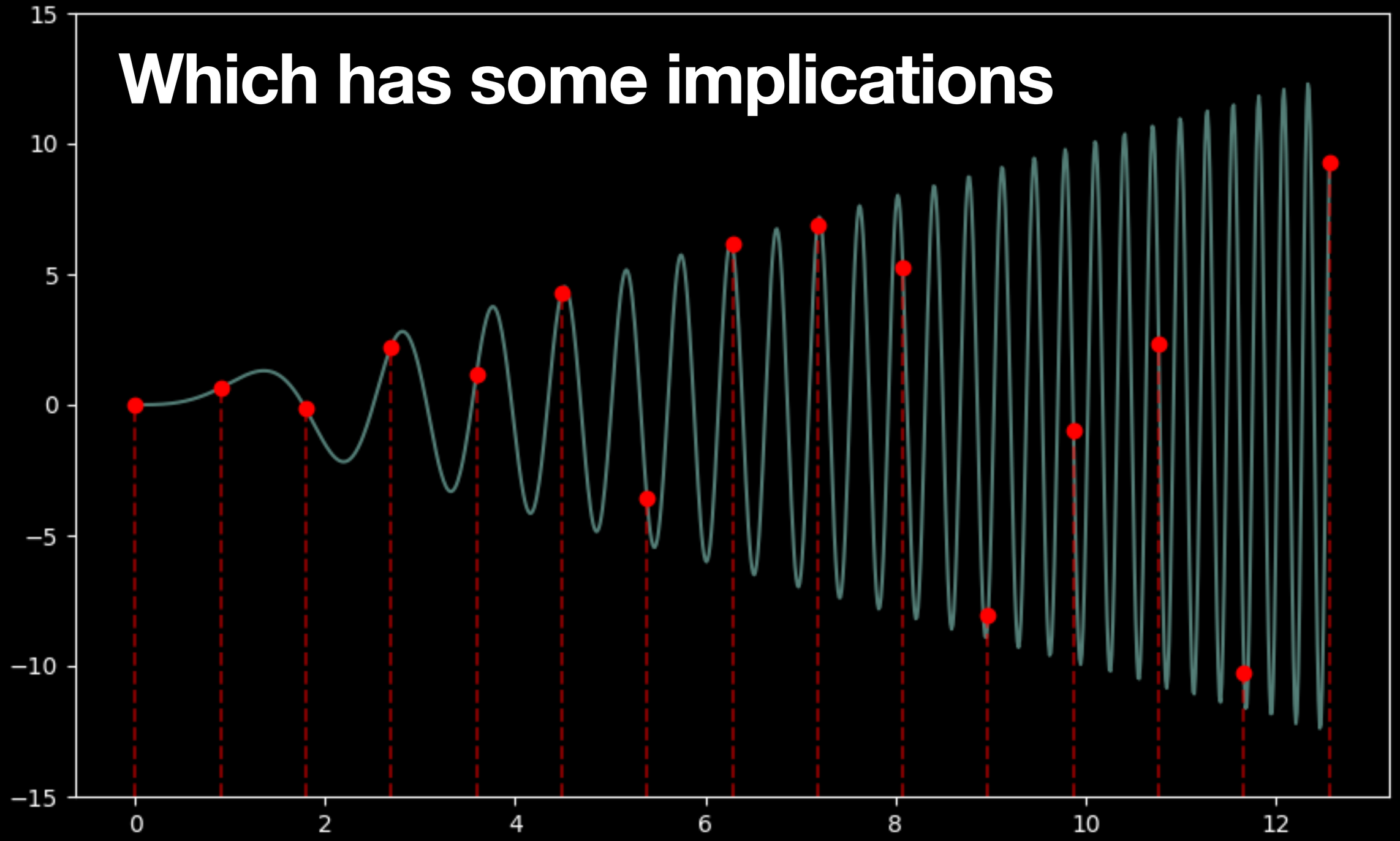
What about in between the pixels?

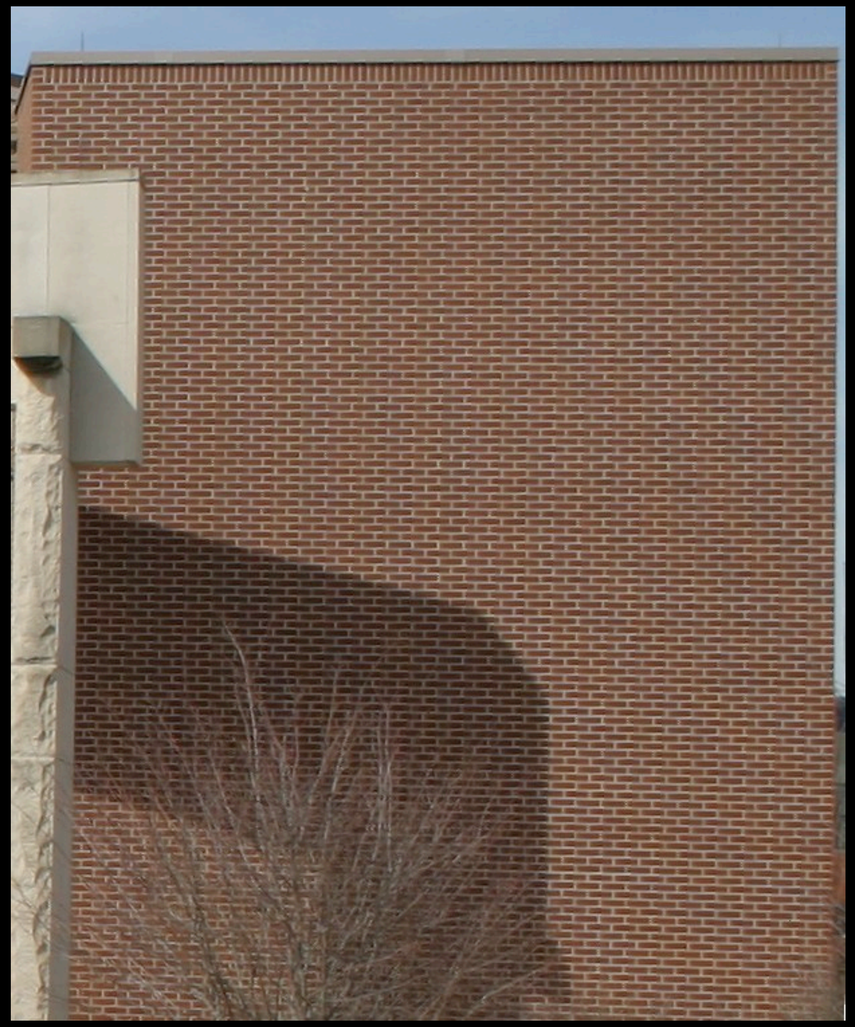
Squares Bilinear Bicubic Lanczos4

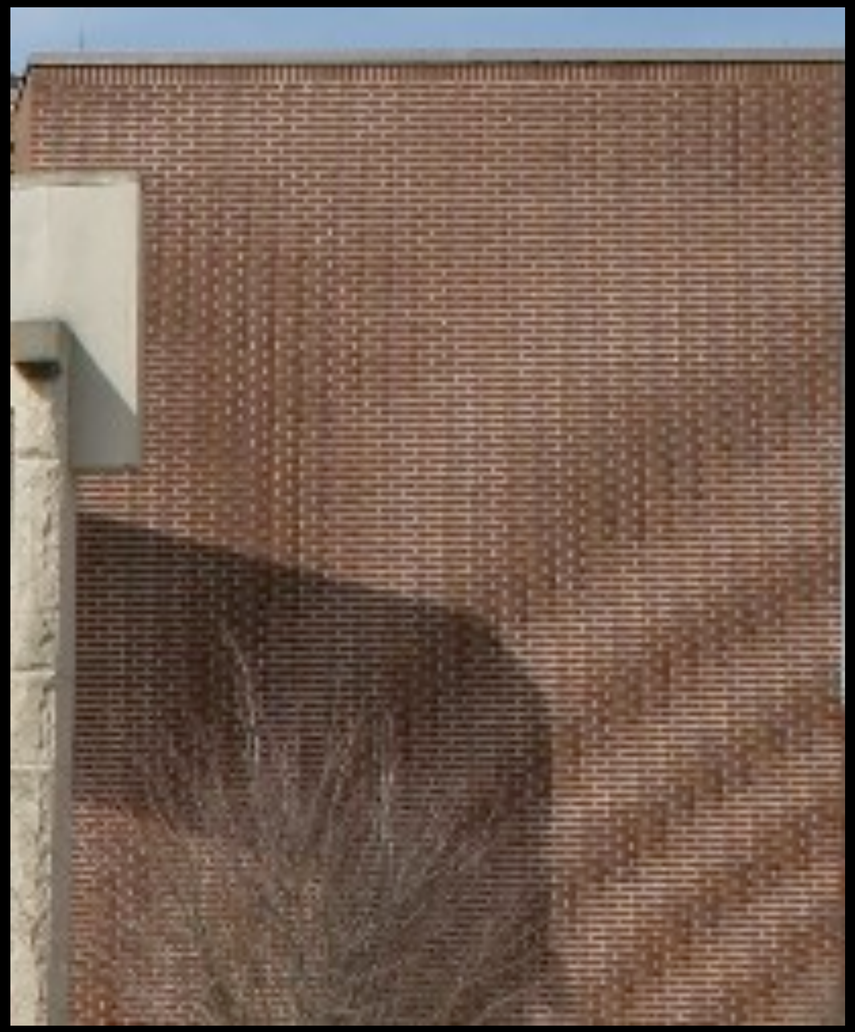
We're sampling a continuous function



Which has some implications







CC BY-SA 3.0, https://commons.wikimedia.org/w/index.php?curid=644927

What is a pixel?

What it’s not…

- A box
- A disk
- A teeny tiny light

It’s a sample…

- It has no dimensions
- It occupies no area
- You can’t see it (it’s infinitely small)
- It’s got a position
- It’s got a value

Pixels Are Not A Square

Why do we care?

- Image Scaling/Resizing
- Image Filtering
- Image Rotation and Transformation
- Frequency Domain Processing

What is a Pixel (part 2)

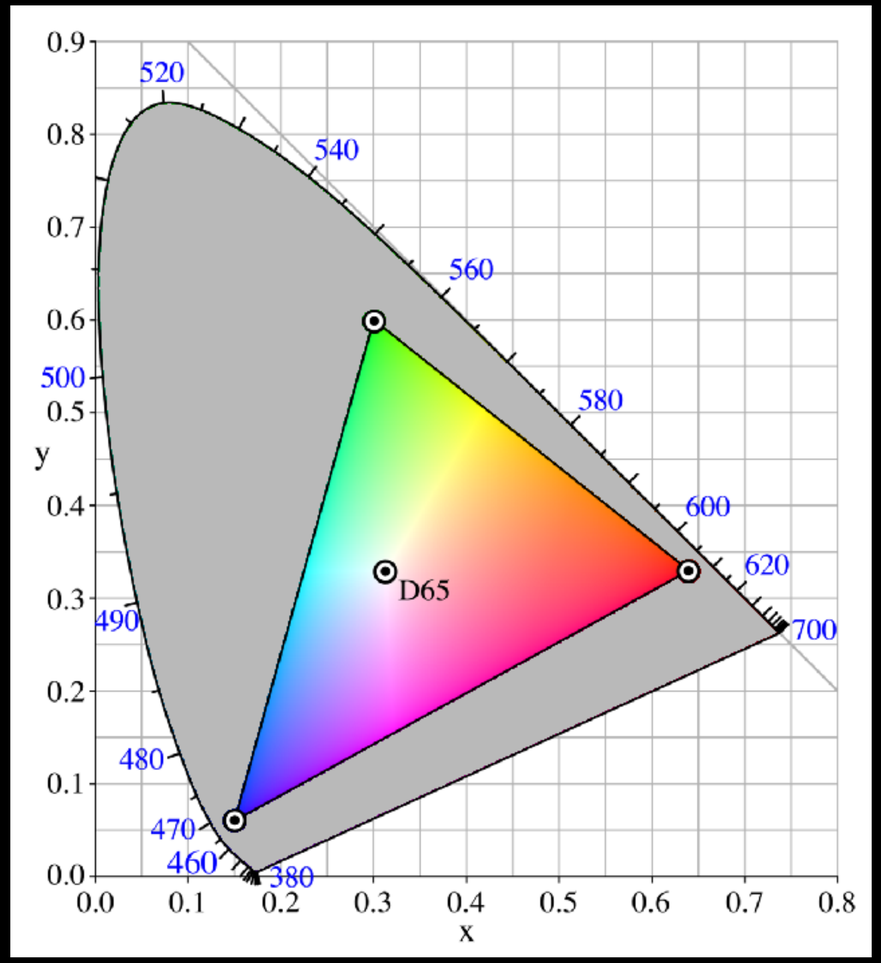
What’s the value of a pixel?

Red, Green, Blue (RGB)

R: 0…255 G: 0…255 B: 0…255

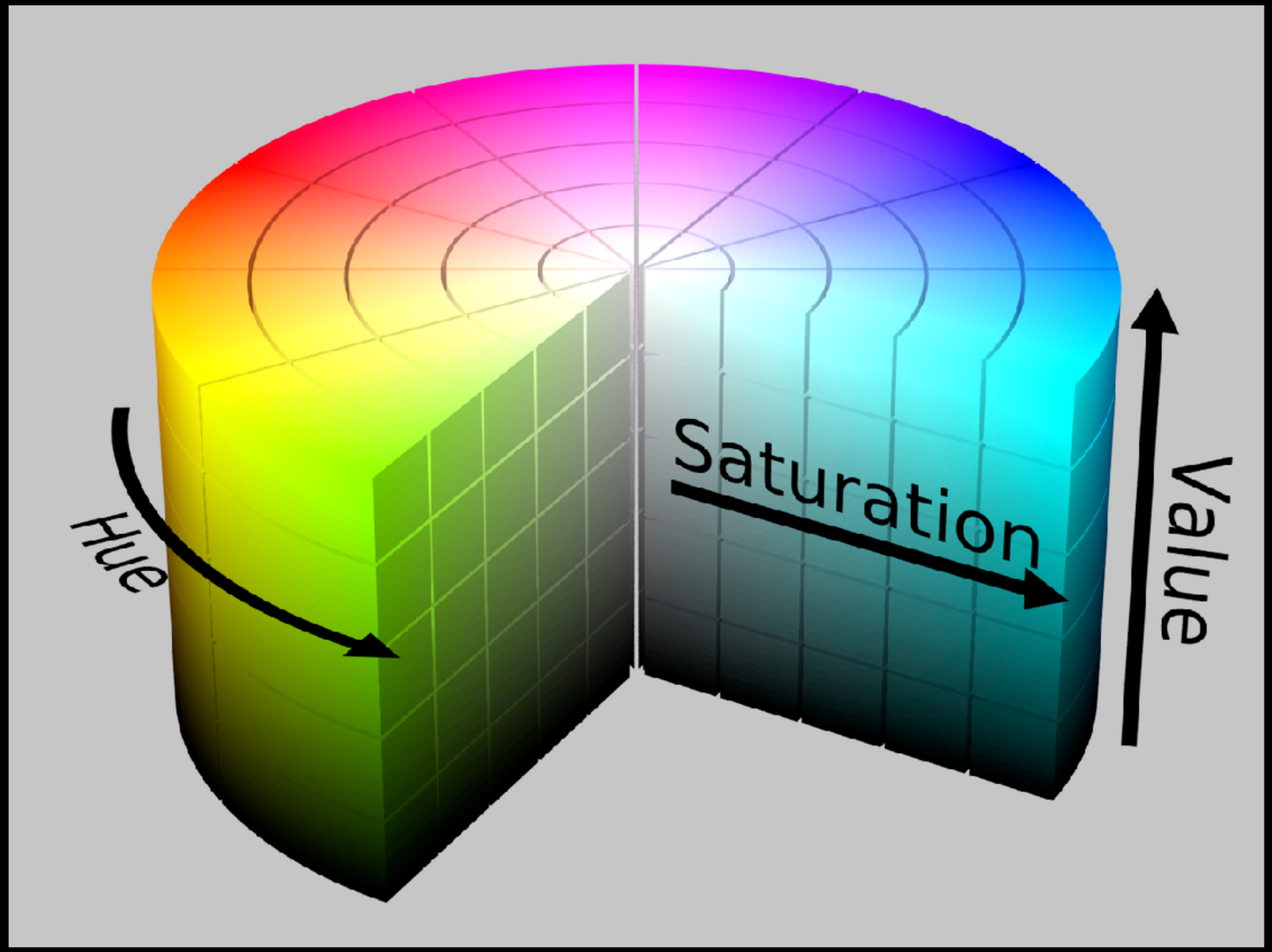
16777216 colors!

Red, Green, Blue (RGB)



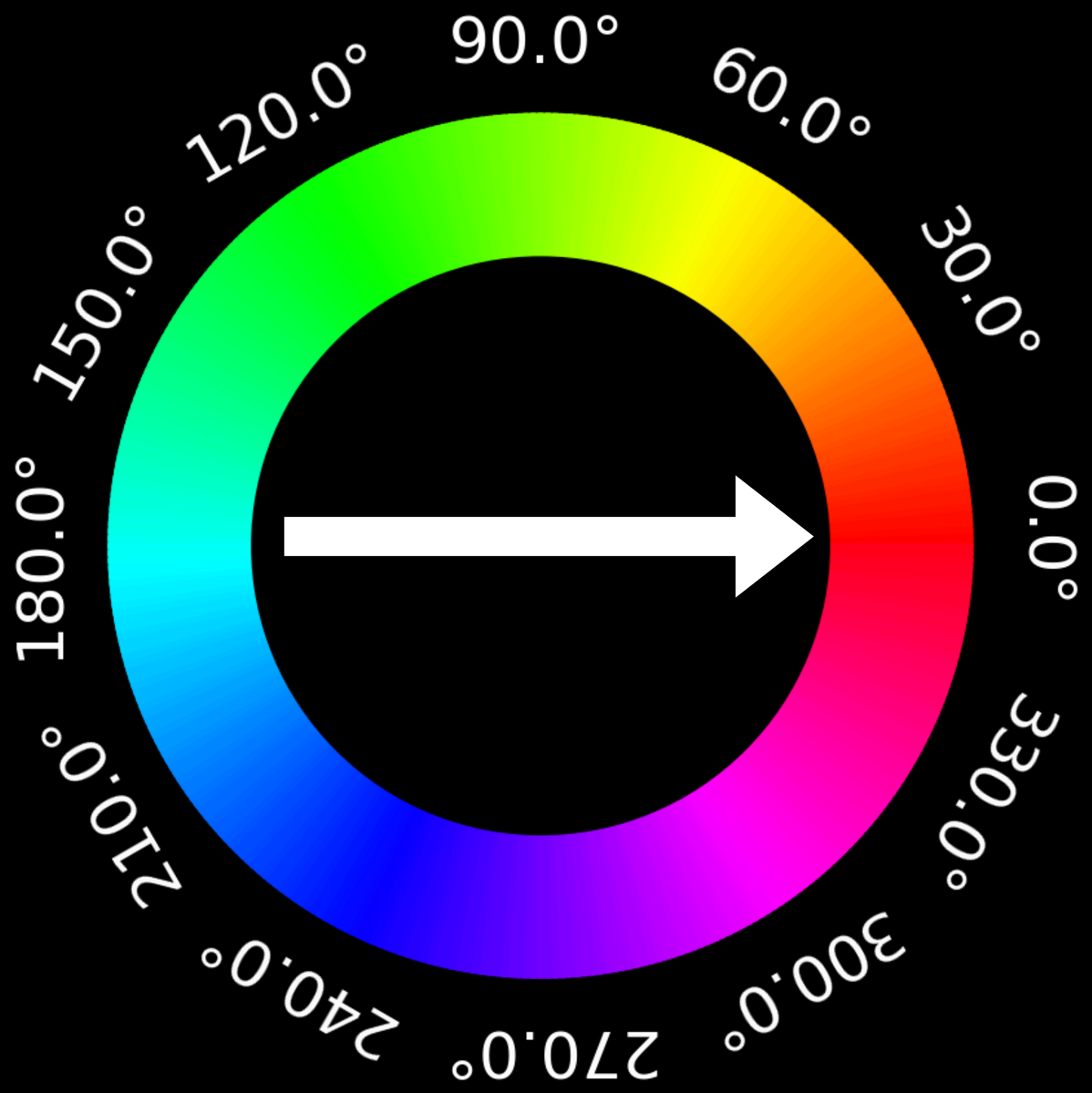
By The original uploader was Dicklyon at English Wikipedia. - Transferred from en.wikipedia to Commons by aboalbiss., CC BY-SA 3.0, https://commons.wikimedia.org/w/index.php?curid=8403395

Hue, Saturation, Value (HSV/HSB)



By HSV_color_solid_cylinder.png: SharkDderivative work: SharkD Talk - HSV_color_solid_cylinder.png, CC BY-SA 3.0, https://commons.wikimedia.org/w/index.php?curid=9801673

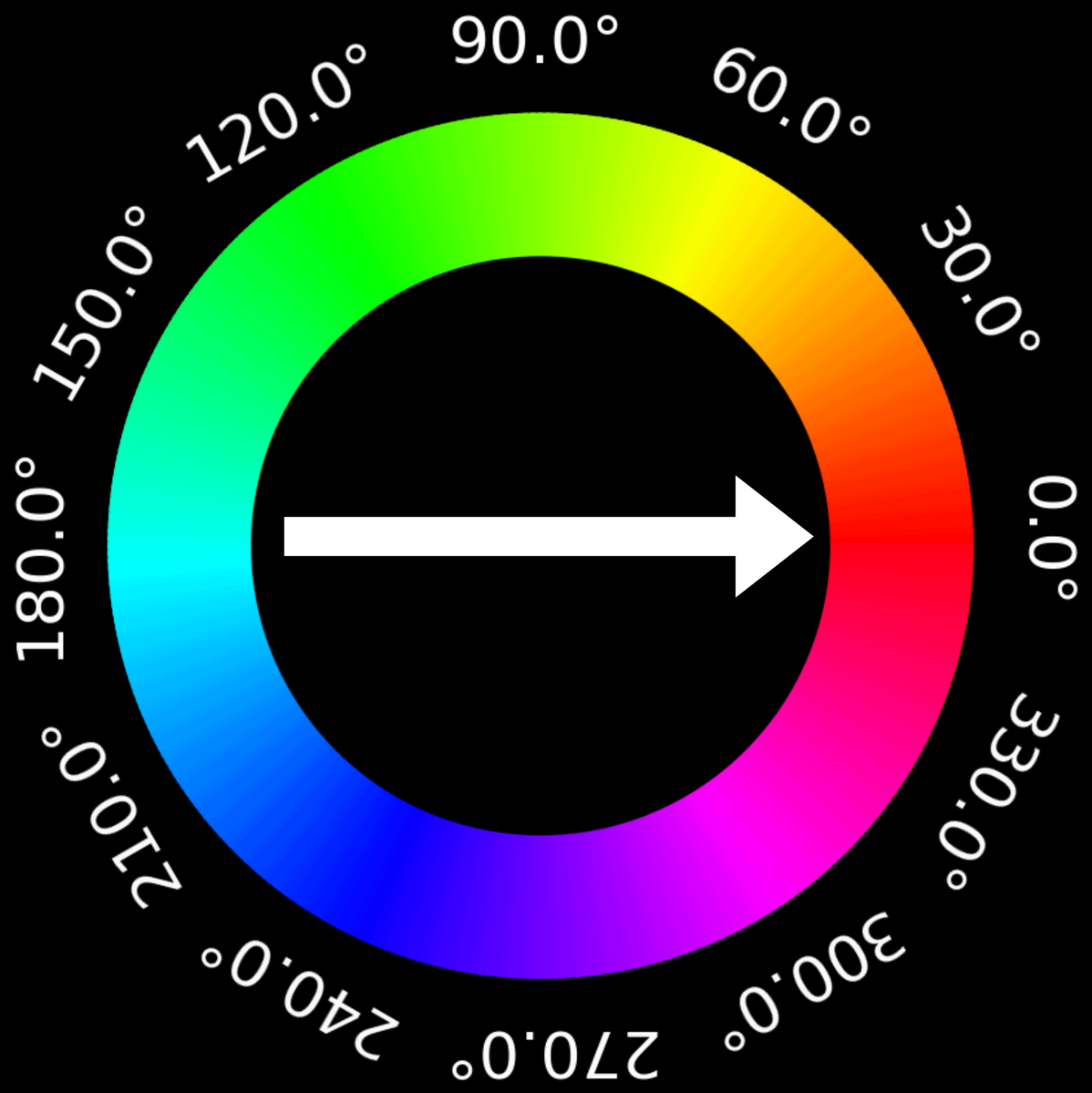
Hue



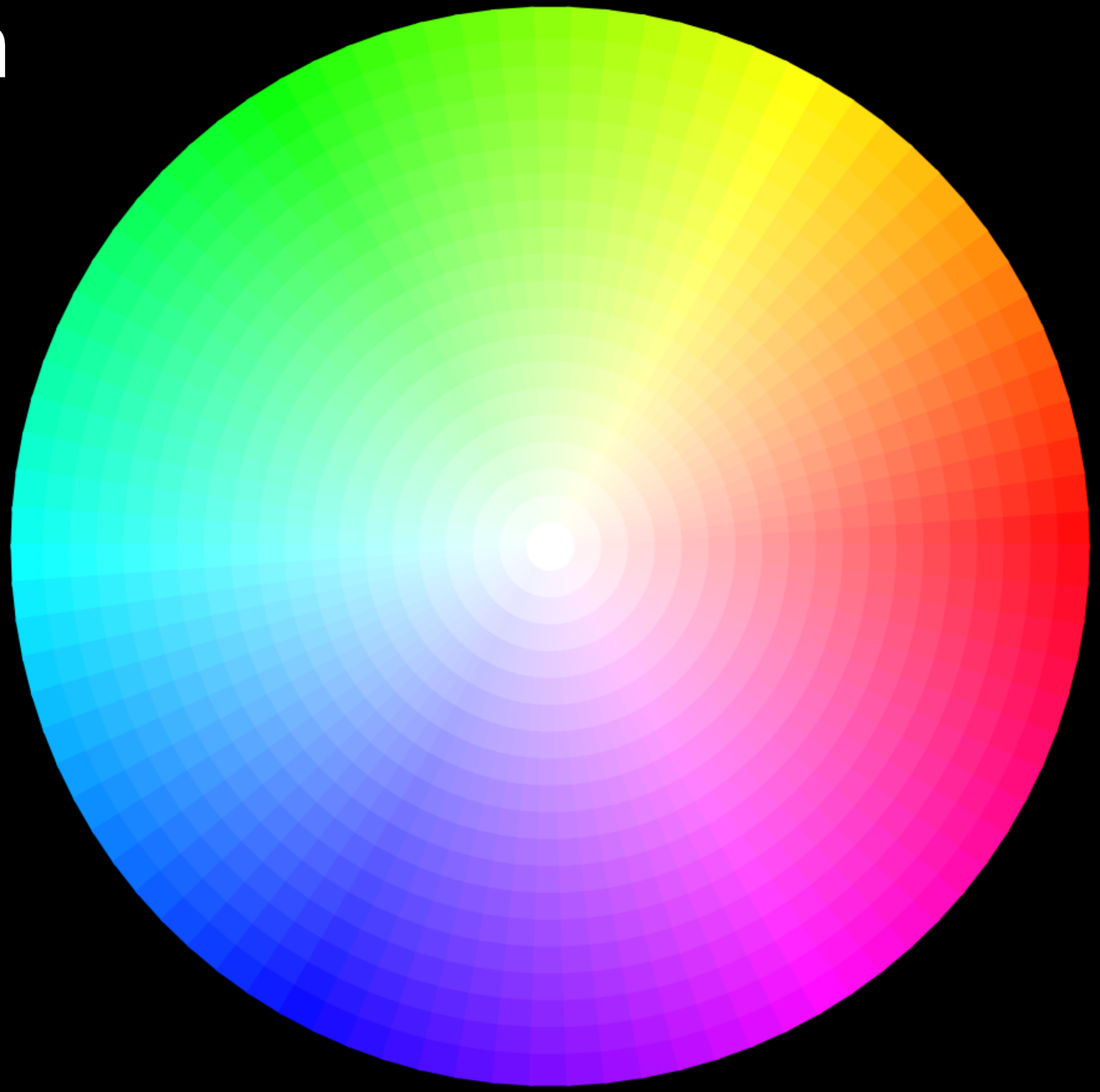
Hue

Hue

Hue



Saturation



Value

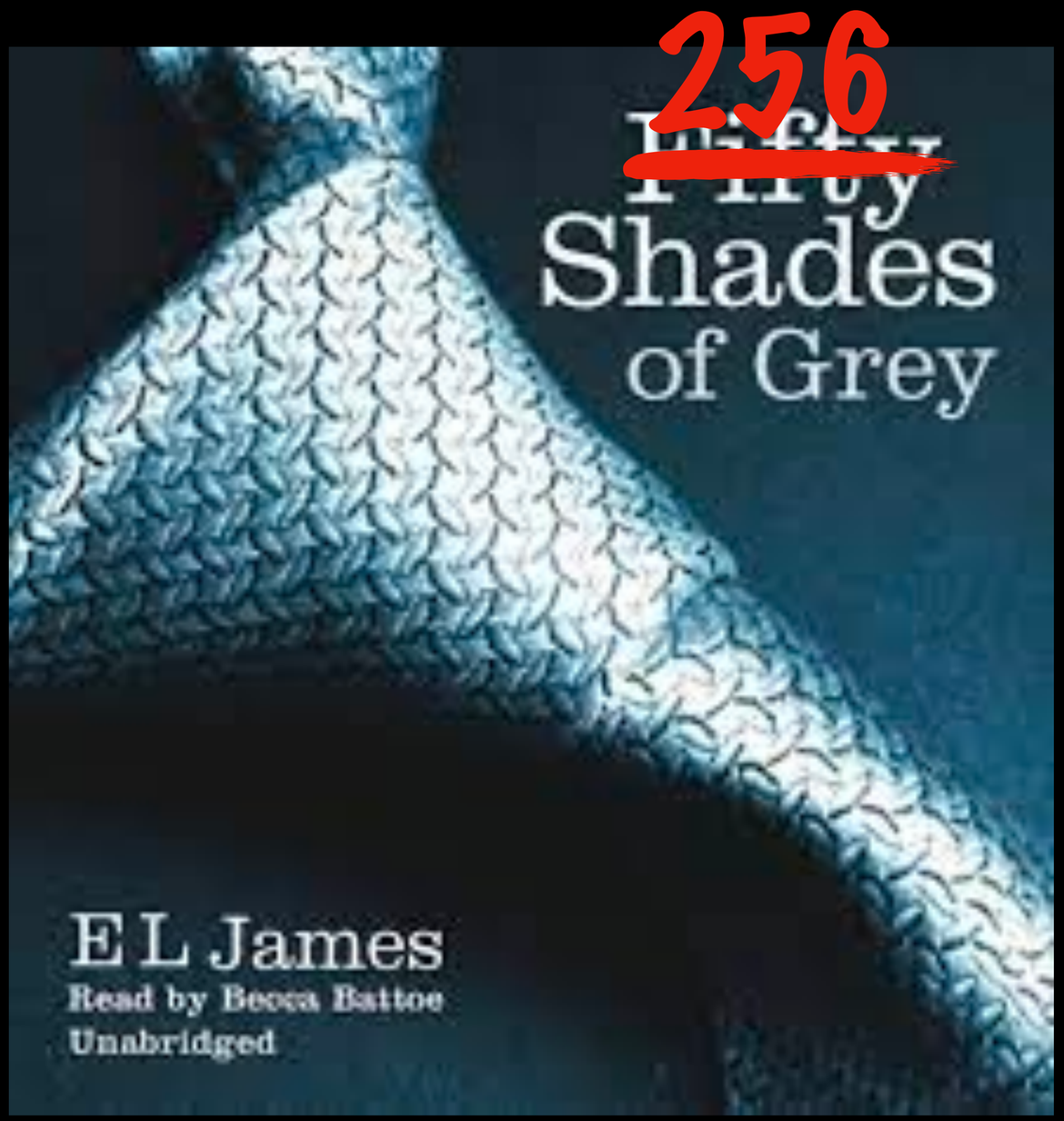
Hue, Saturation, Value (HSV)

Hue

Saturation

Value

Greyscale


256
Fifty
Shades
of Grey
E L James
Read by Becca Battoe
Unabridged

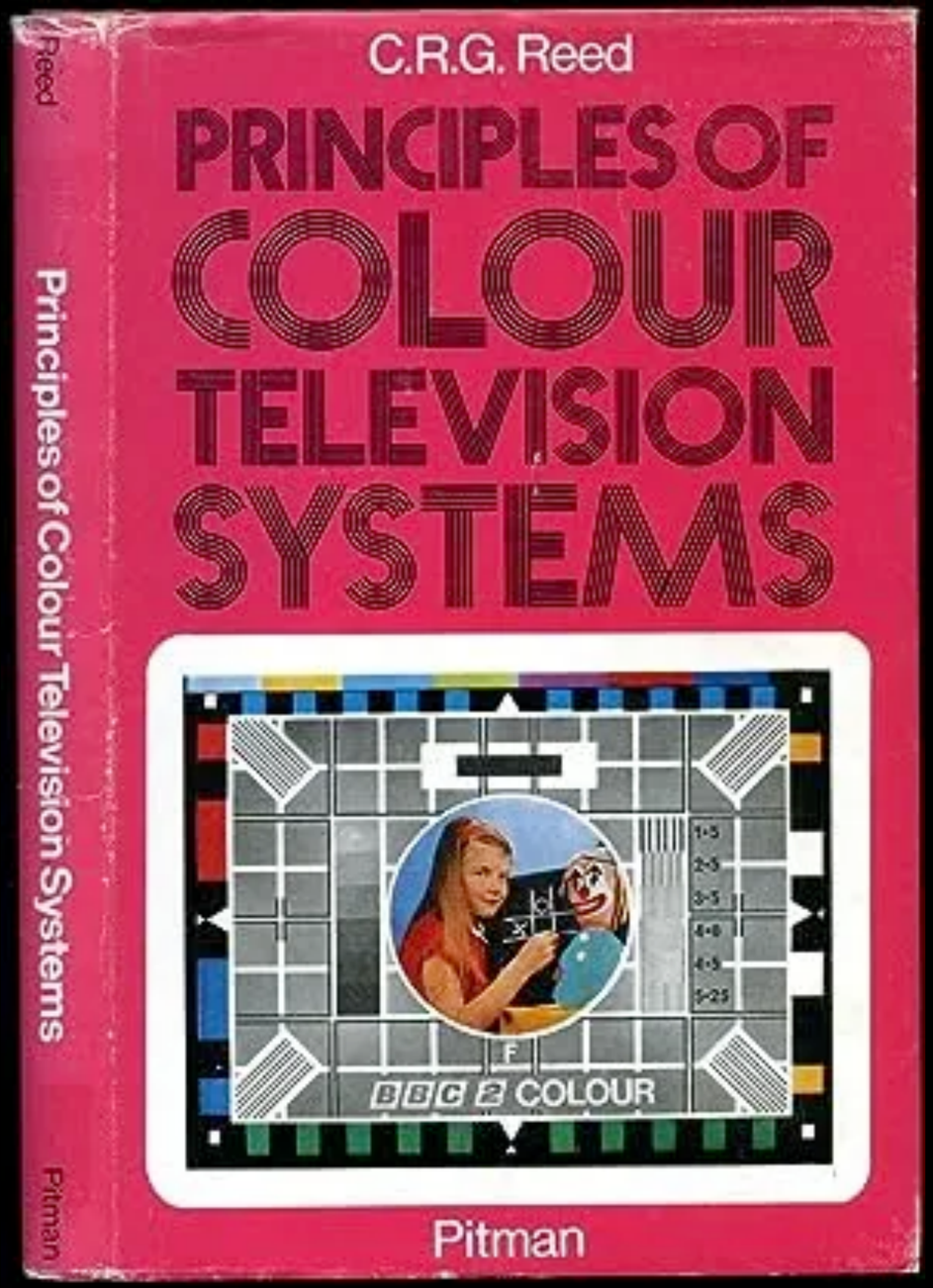
Greyscale

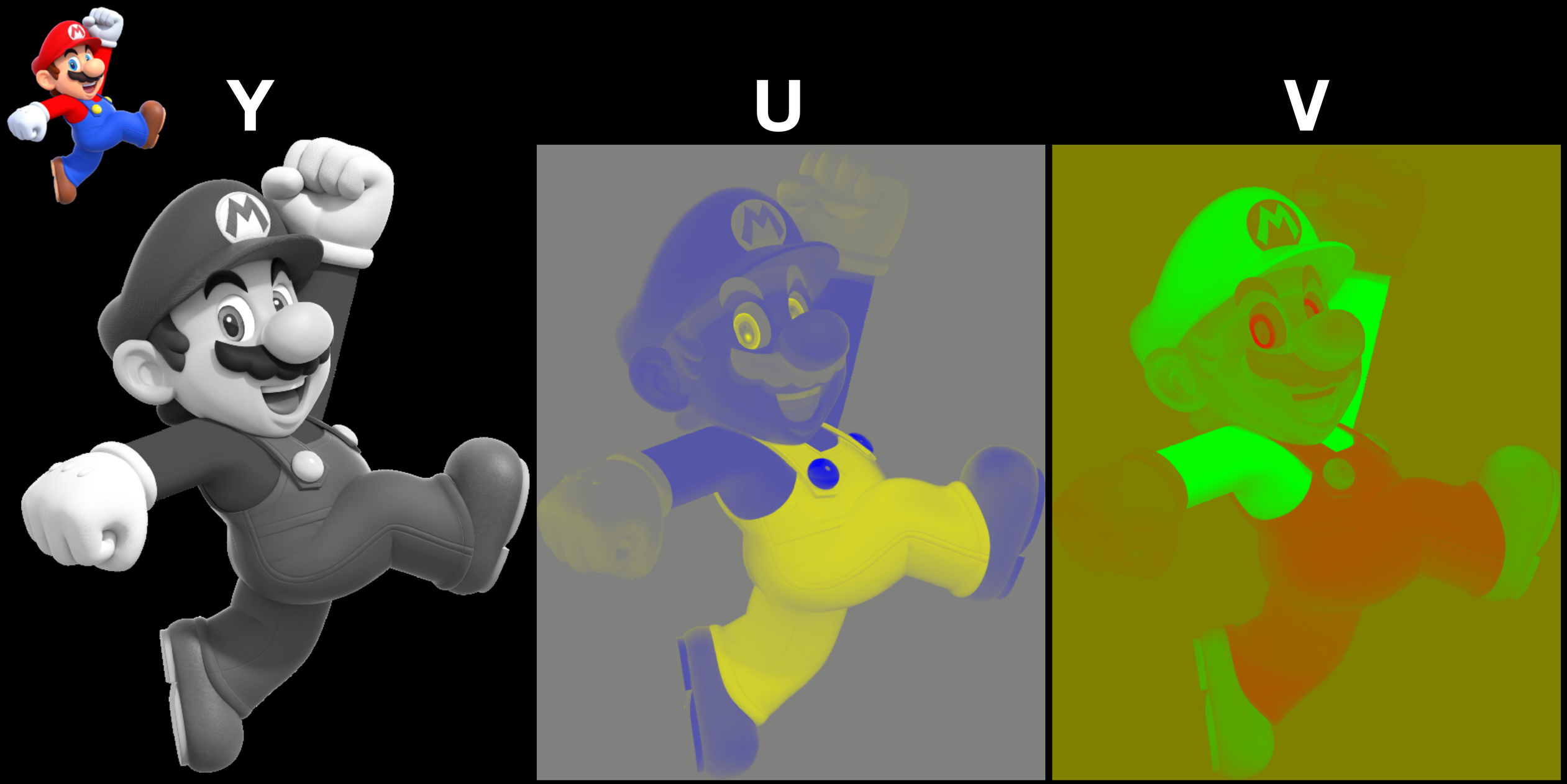
$(R + G + B)/3$

$(0.21 \times R) + (0.72 \times G) + (0.07 \times B)$

YUV

- **Y** - Luminance
- **U** - Chrominance (blue)
- **V** - Chrominance (red)

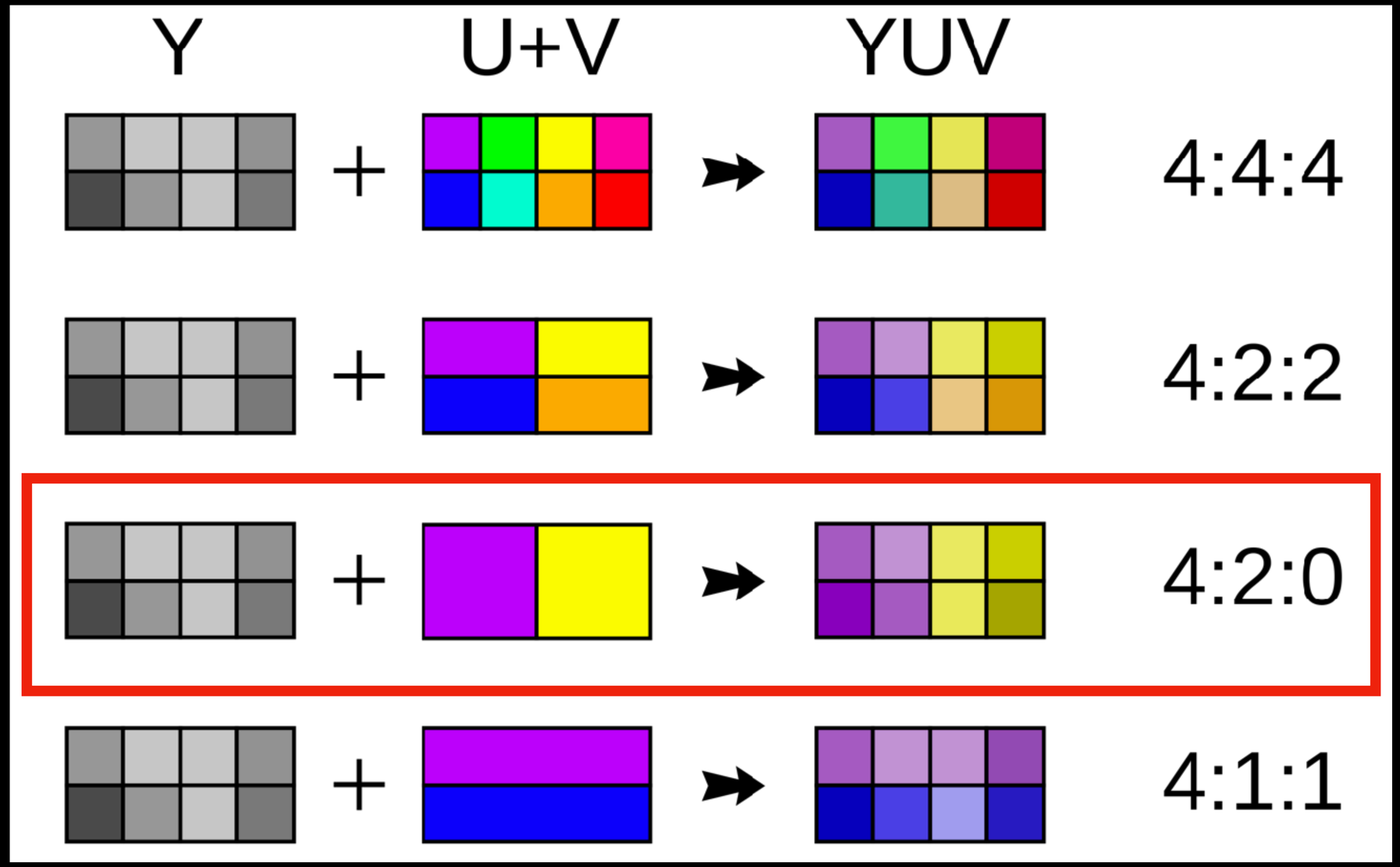



Y
U
V

By Stevo-88 - Own work, Public Domain, https://commons.wikimedia.org/w/index.php?curid=10942363

Why? YUV

Perception is everything



Why YUV

Subsampling

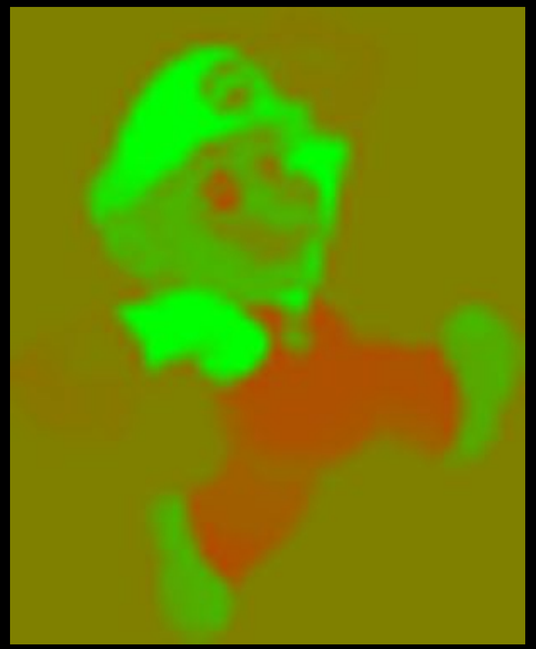
1x1 2x2 4x4 8x8 16x16 32x32

Why YUV

8x8 UV Downsampling

Y

U



V

Result

Le Voyage dans la Lune (A Trip to the Moon) 1902

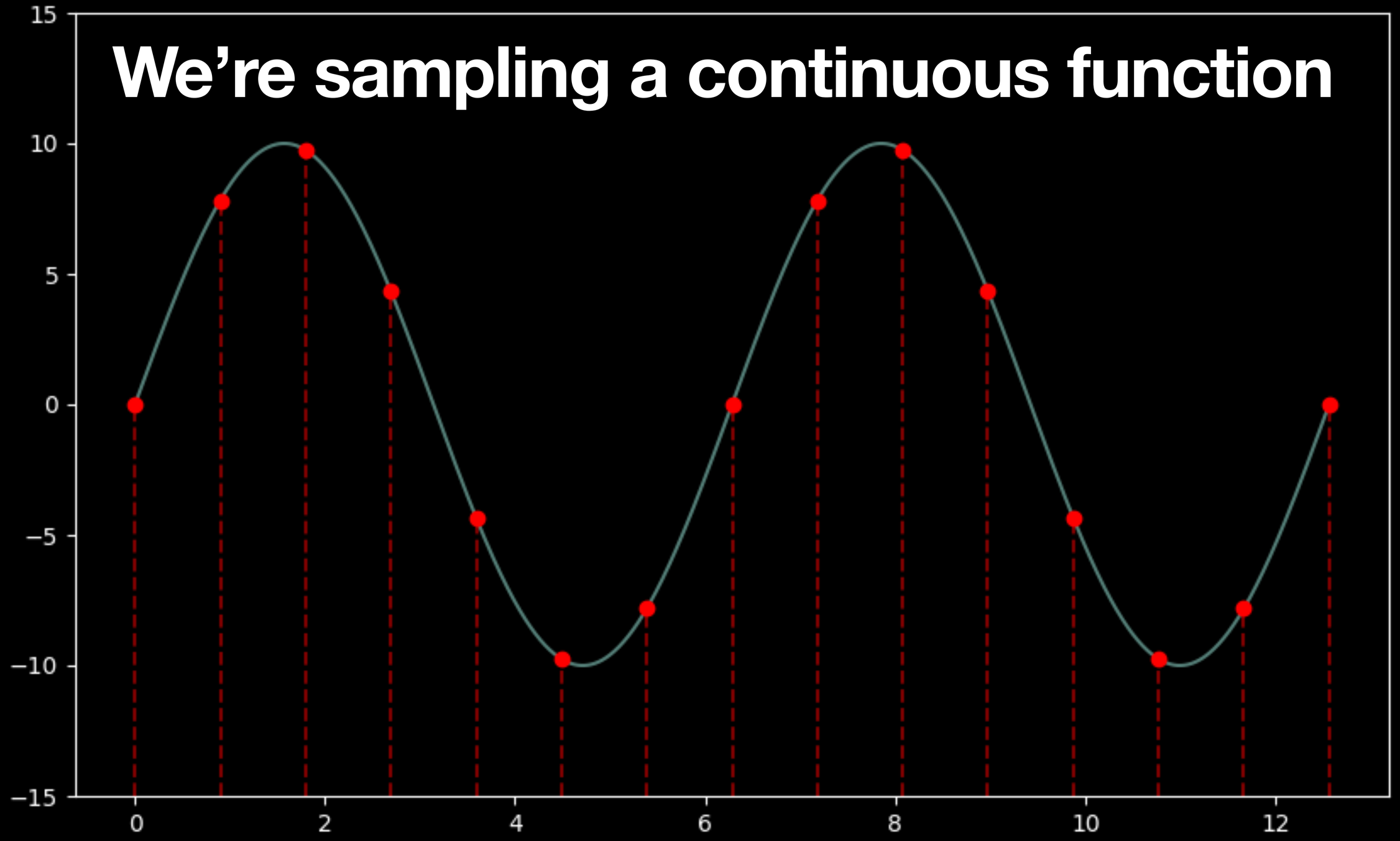
CMYK

You probably shouldn’t need to worry about this…

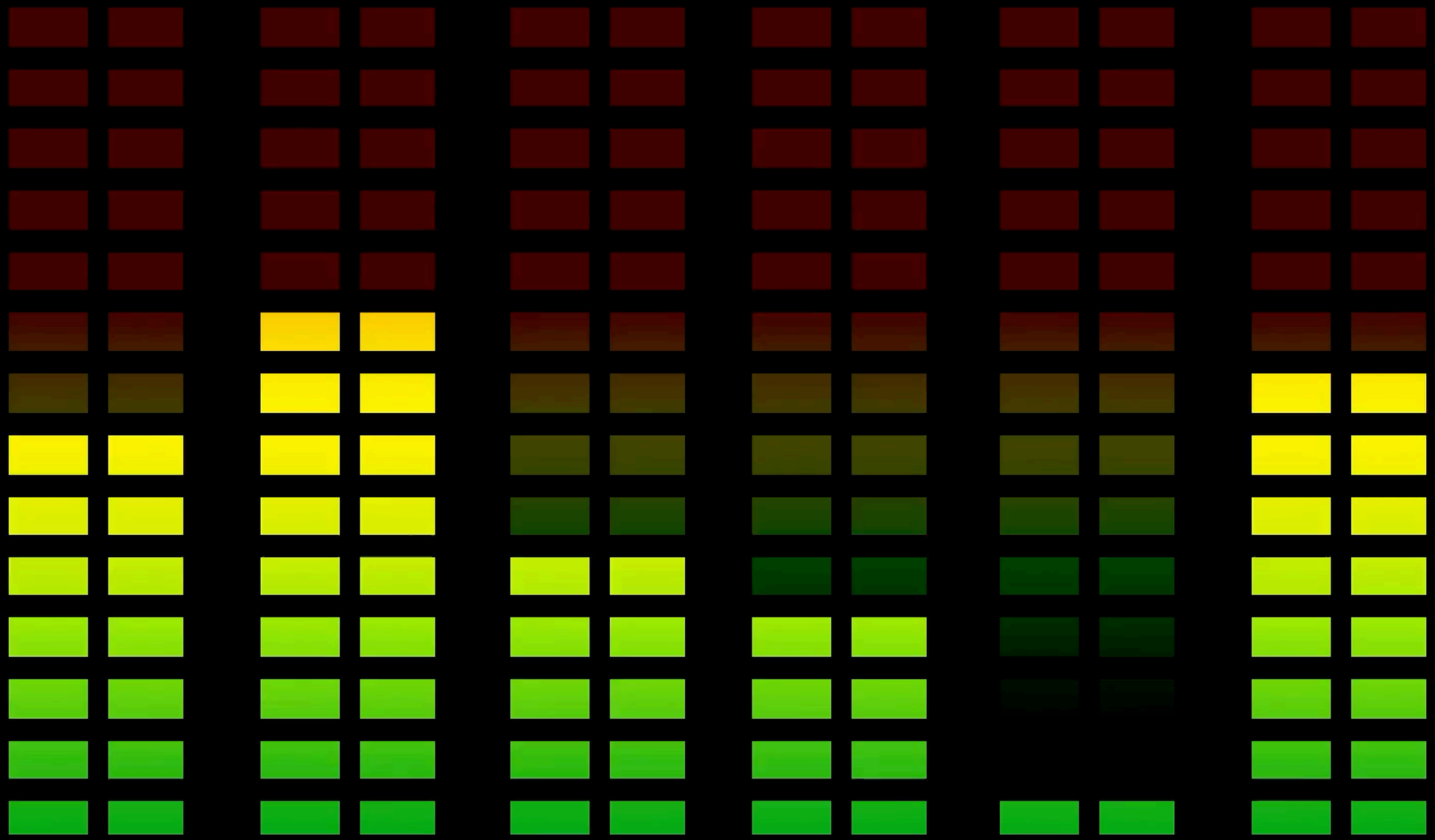
There are a whole bunch more…

- Lab (or CIELAB)
- LCH (or CIELCH)
- XYZ (or CIEXYZ)
- Adobe RGB
- HSL (Hue, Saturation, Lightness)
- ProPhoto RGB
- CMY (Cyan, Magenta, Yellow)
- xyY
- ACES (Academy Color Encoding System)
- IPT (Izzat, Pan, Tingle)
- etc…

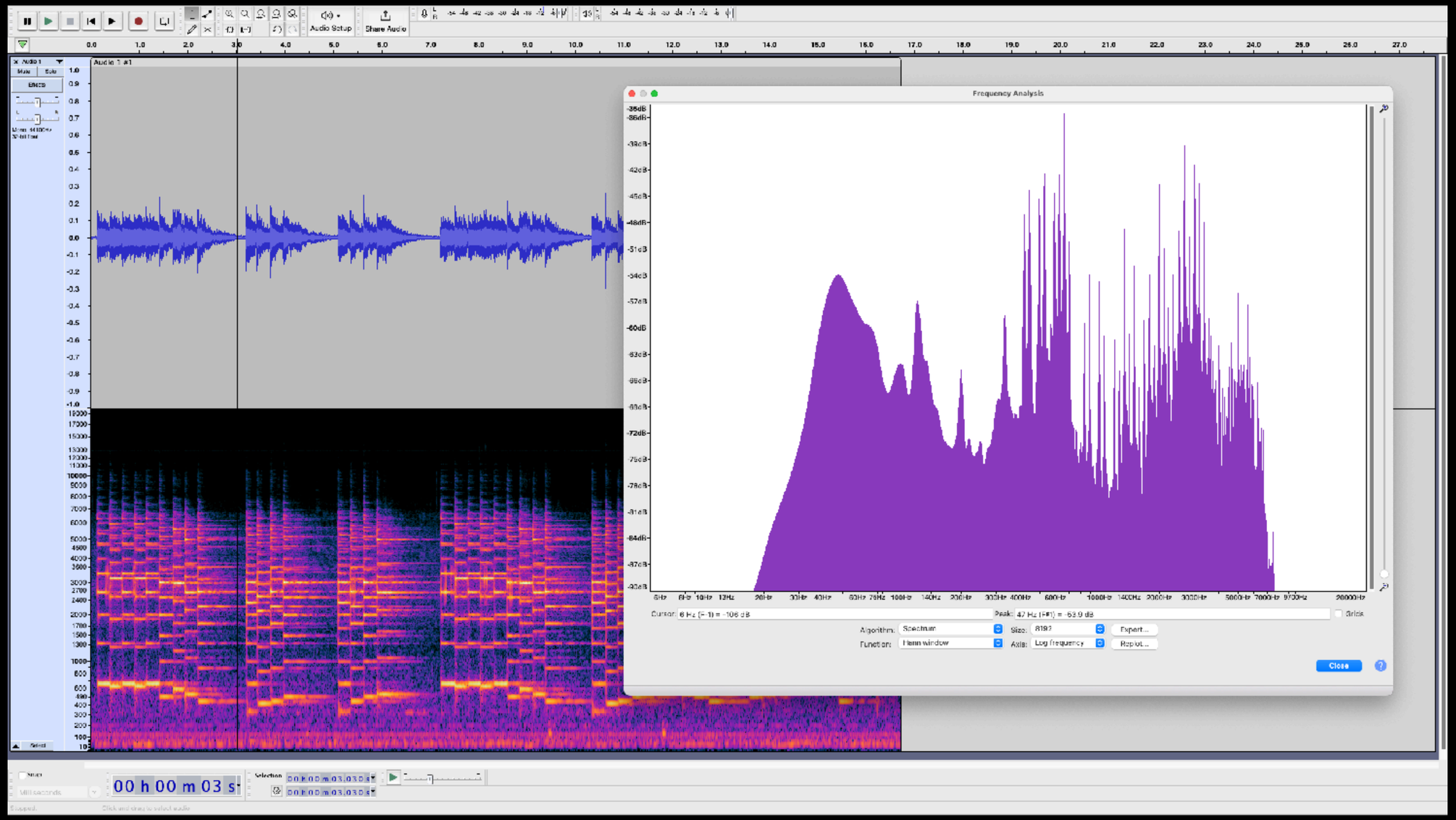
We're sampling a continuous function



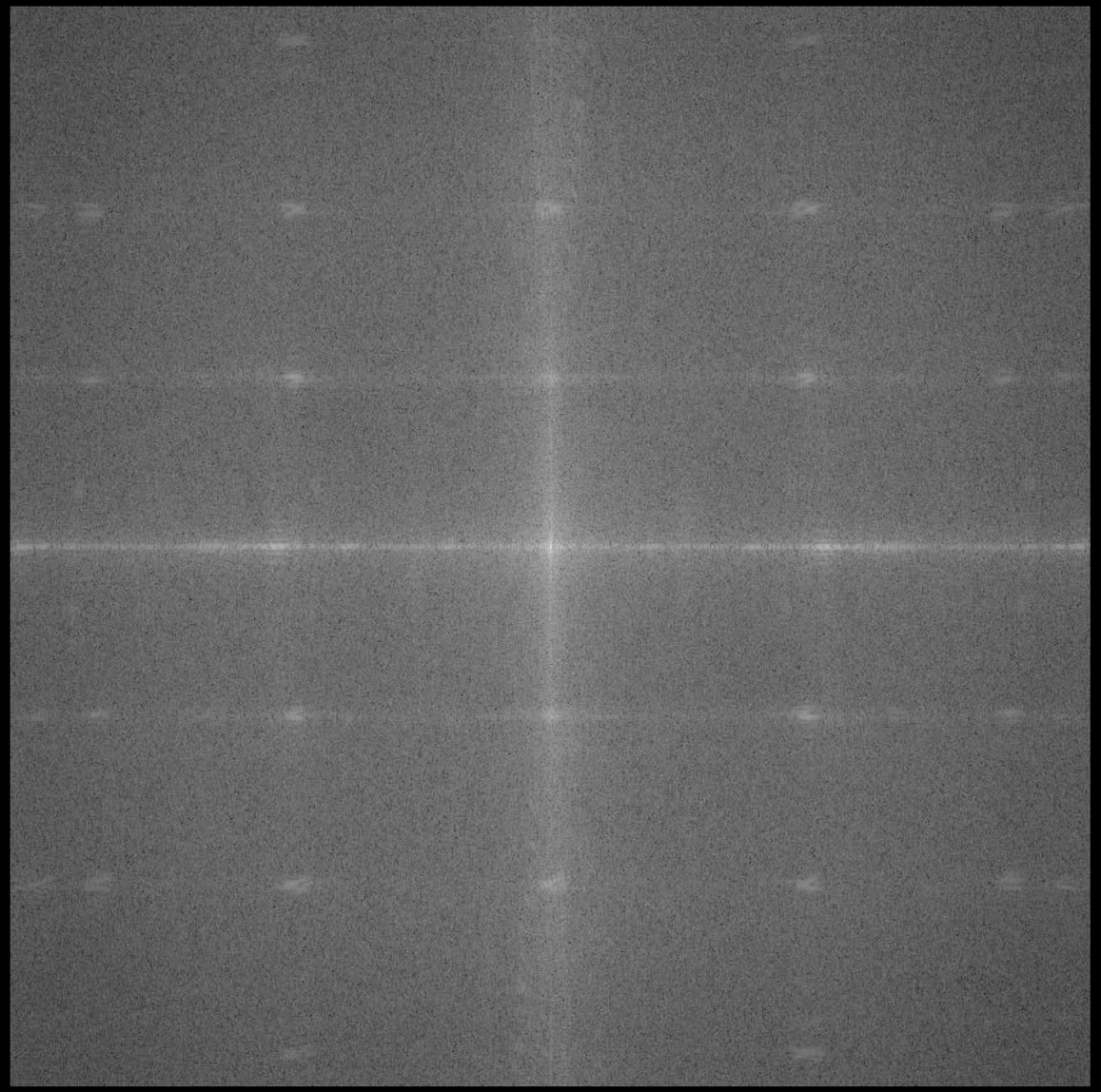
Frequency Domain



Frequency Domain

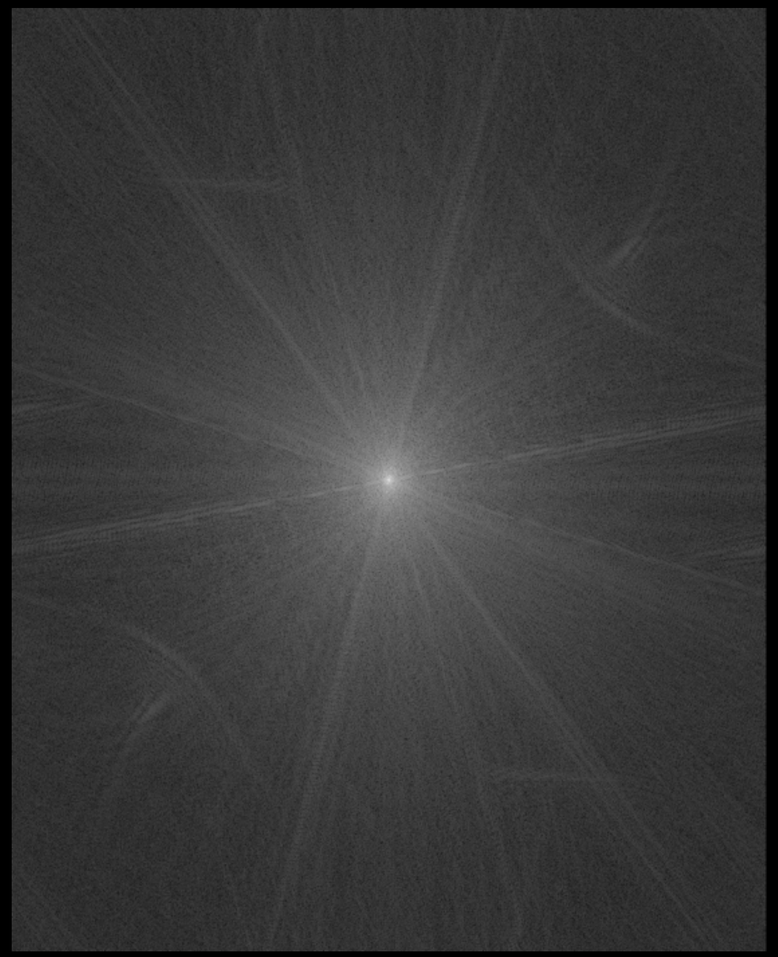


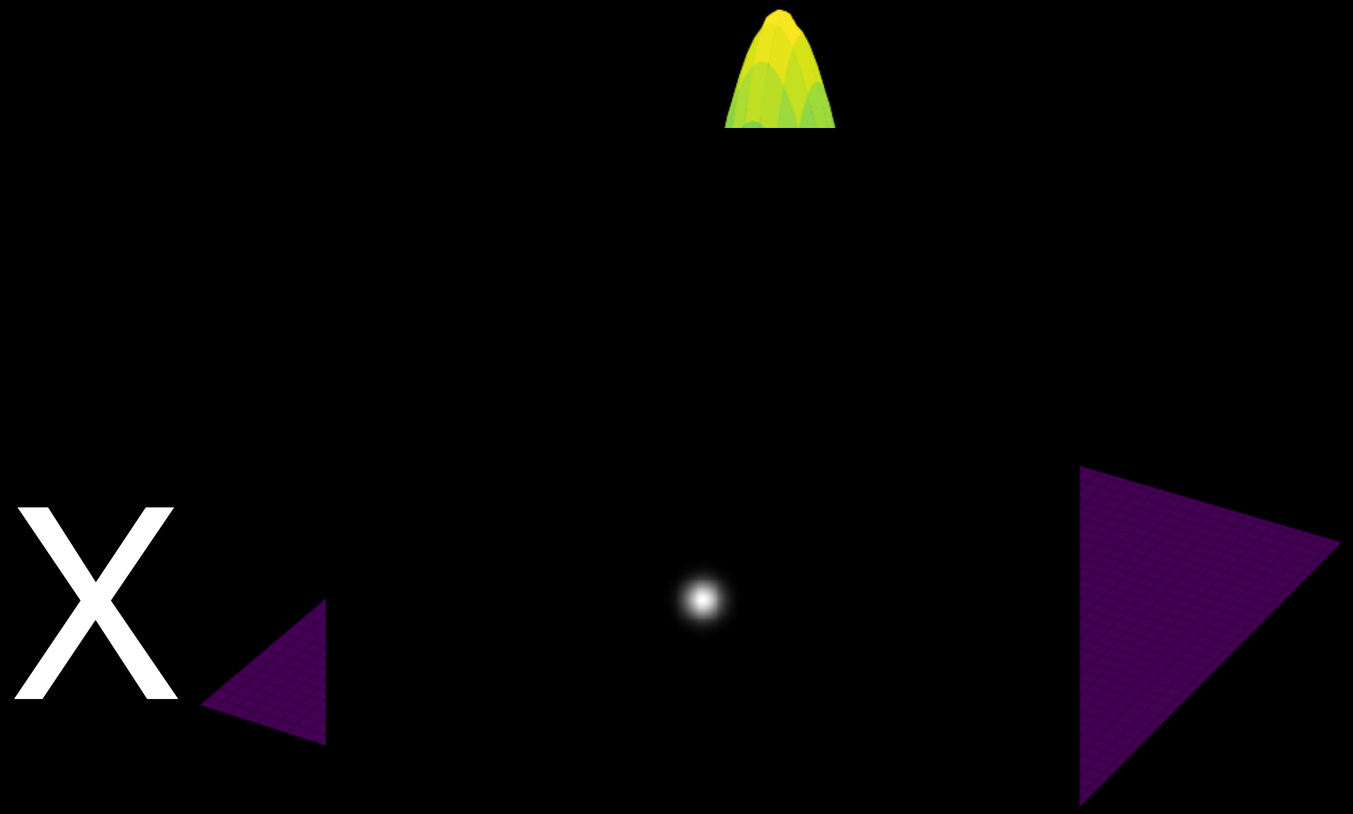
Frequency Domain



Low Pass Filtering

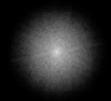
AKA Smoothing/Blurring





Low Pass Filtering

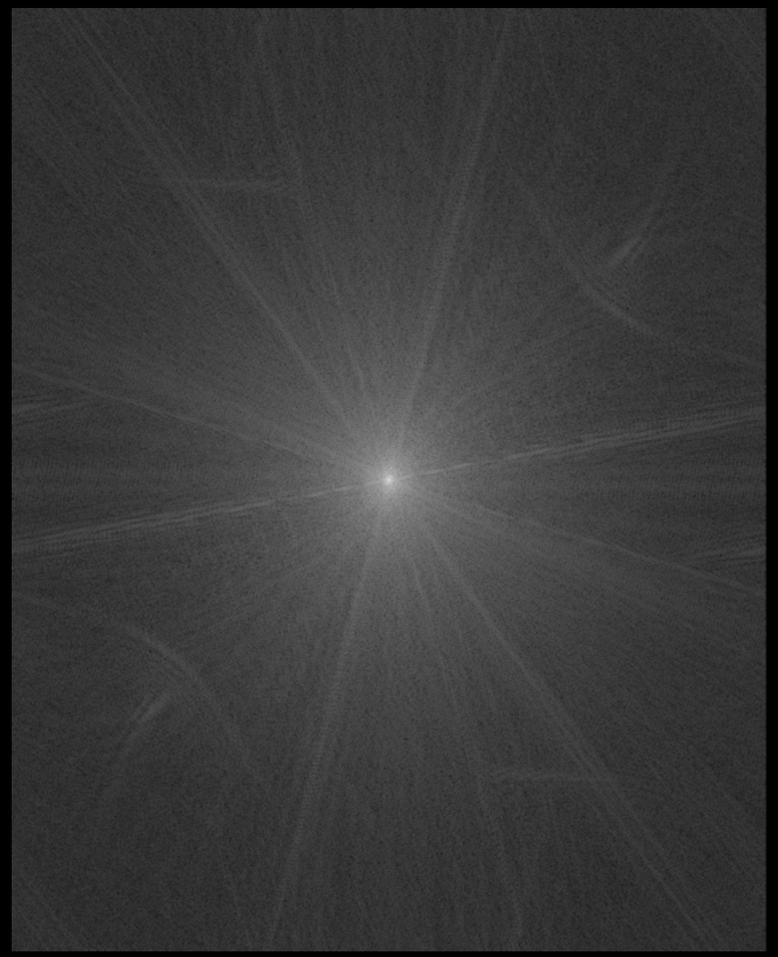
AKA Smoothing/Blurring



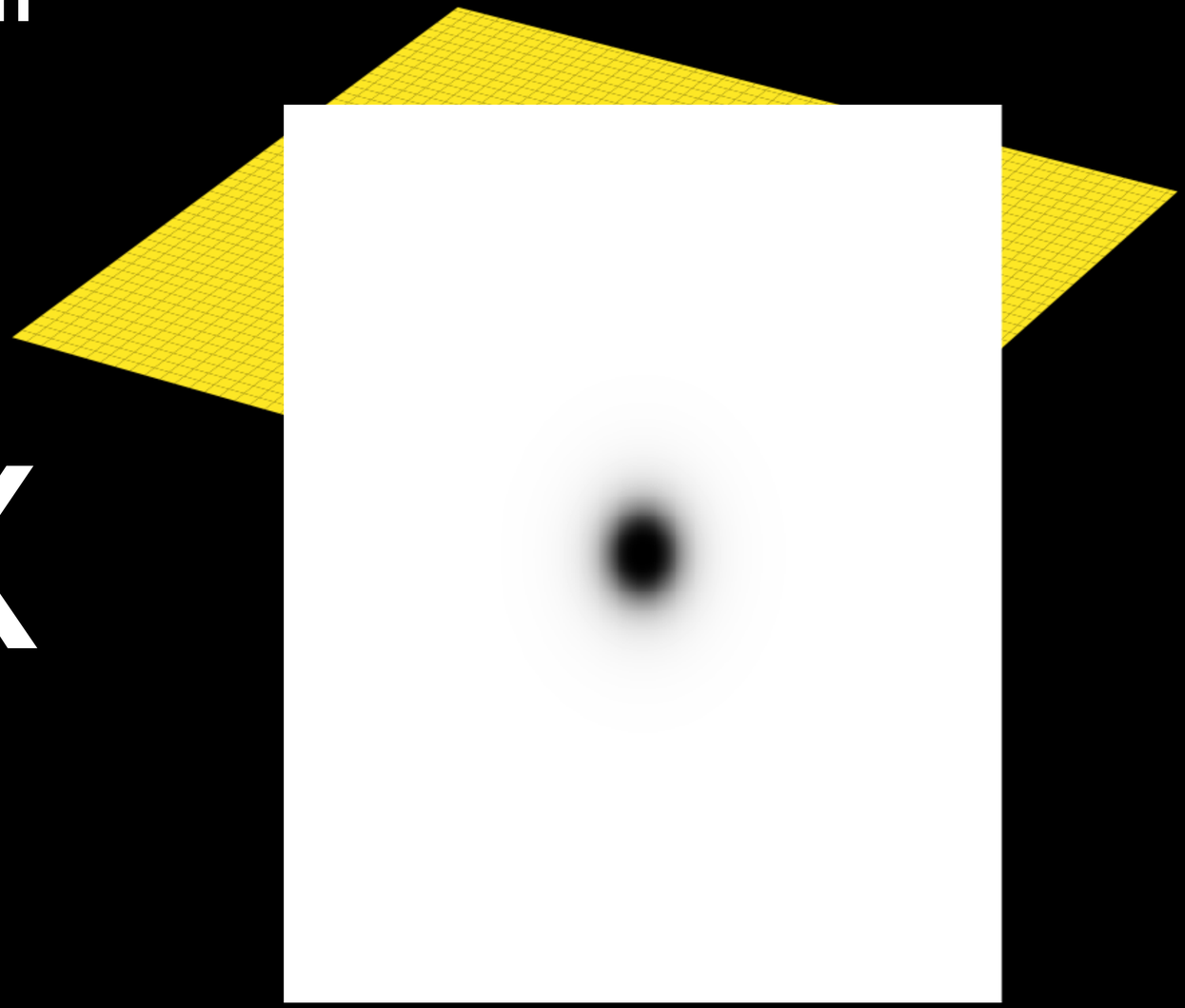
=

High Pass Filtering

AKA Edge Detection/Line Detection

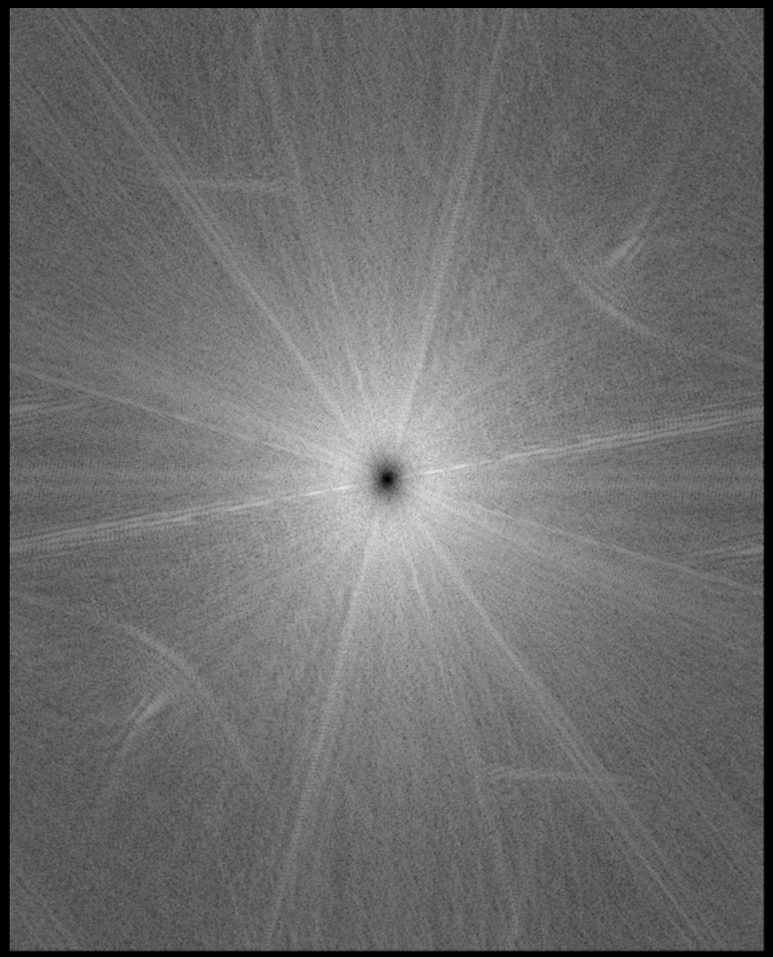


X



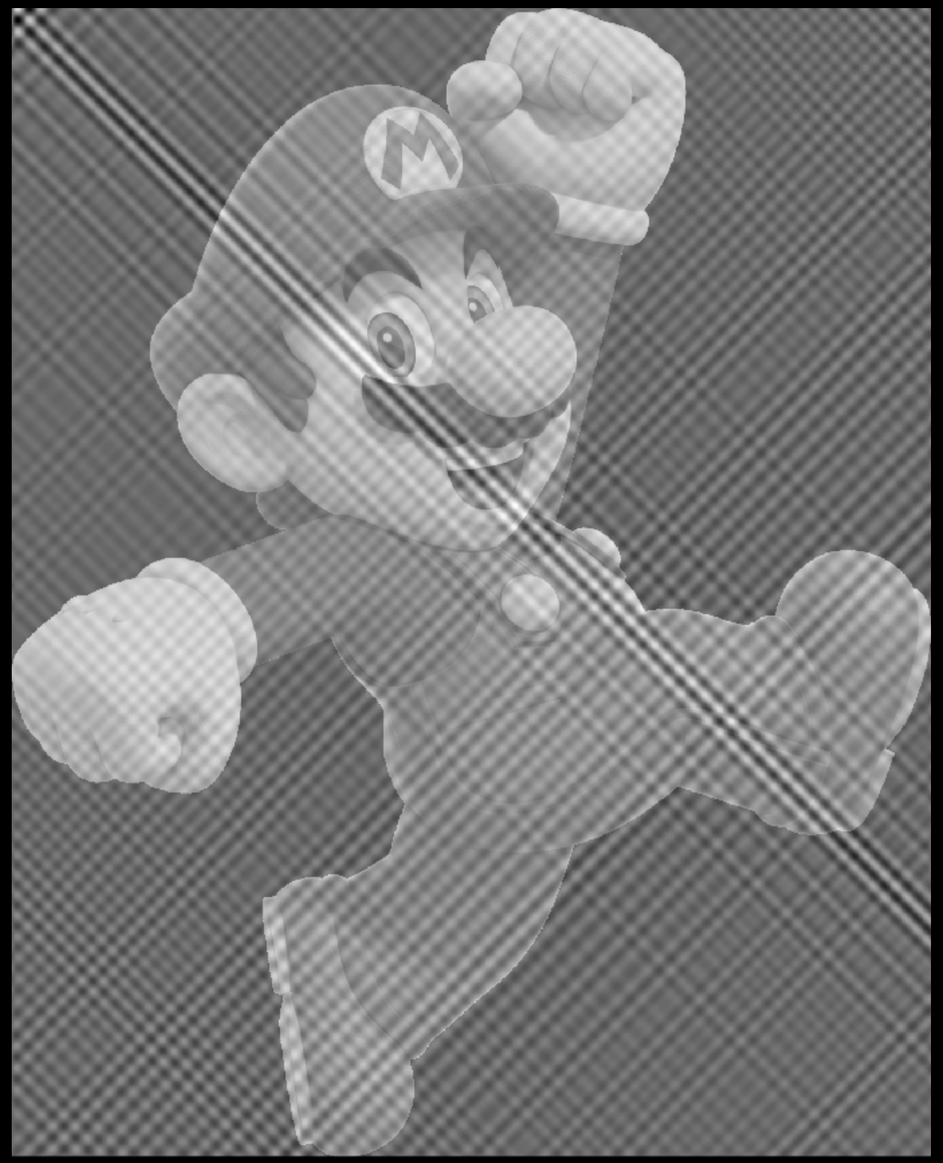
High Pass Filtering

AKA Edge Detection/Line Detection

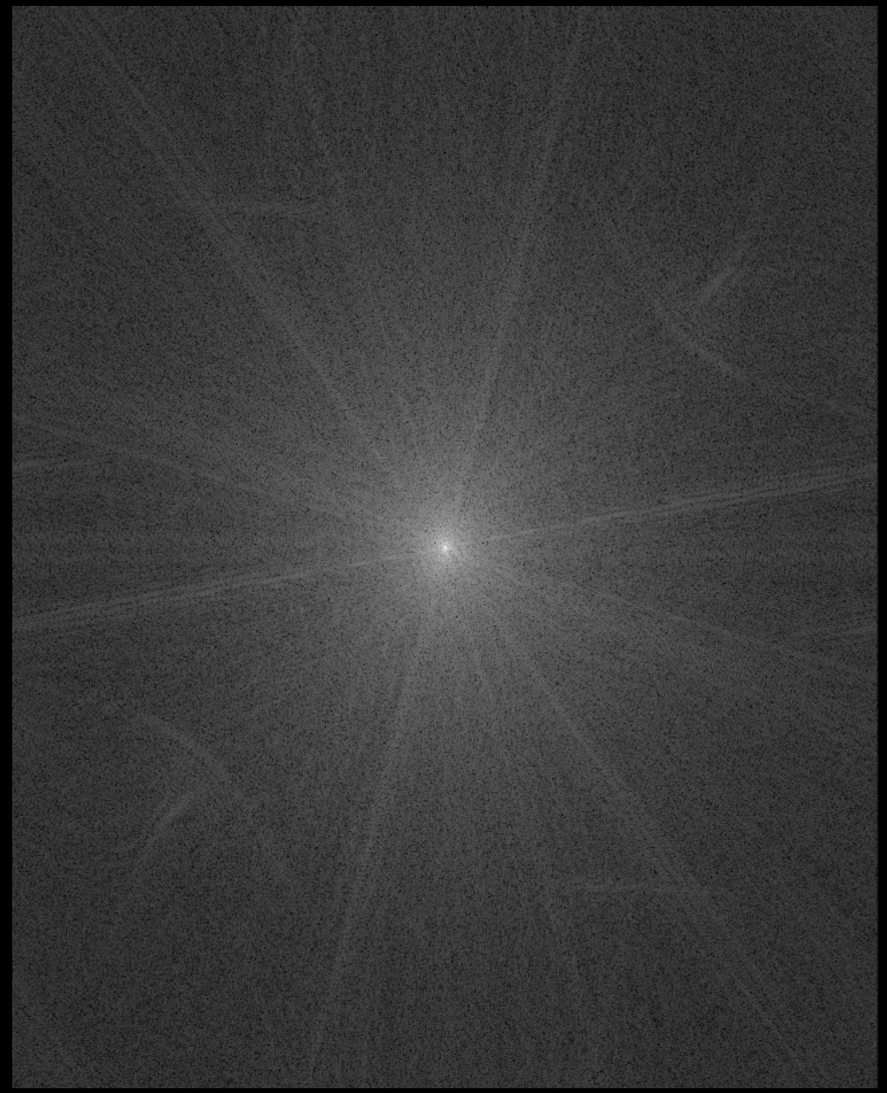


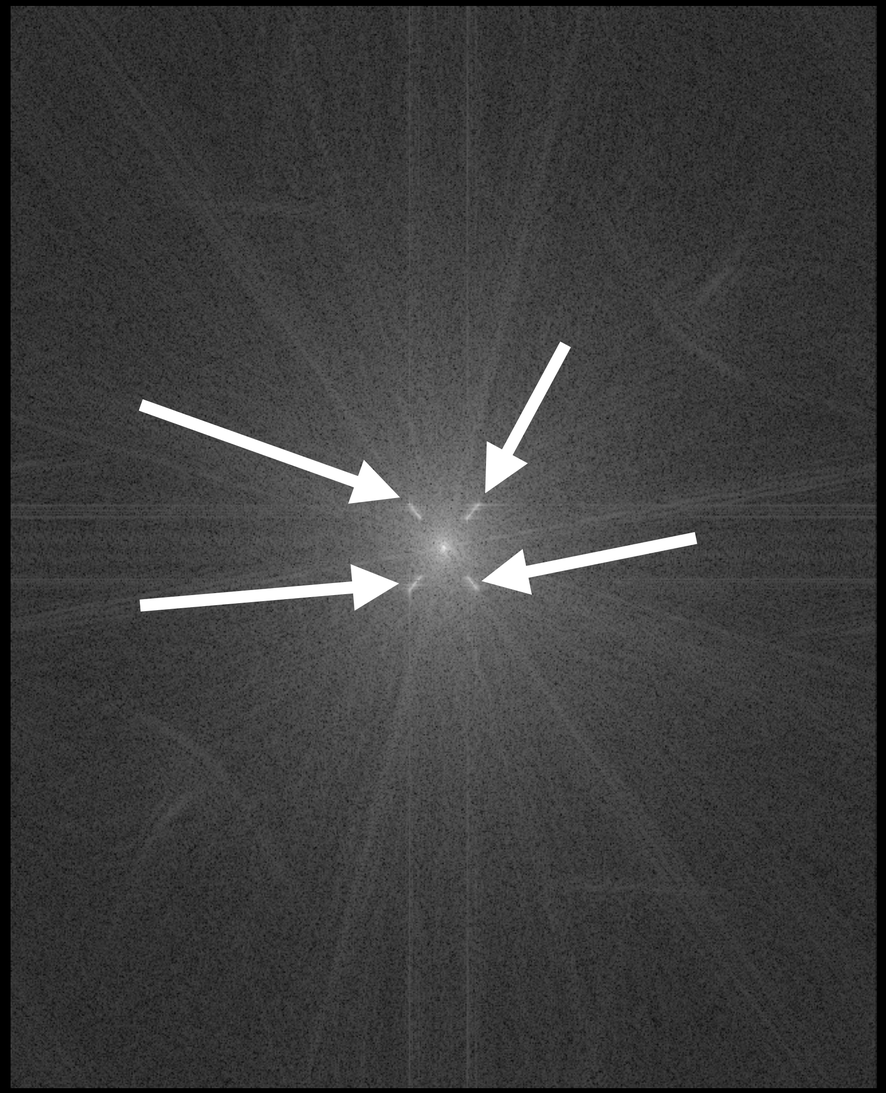
=

Noise Removal

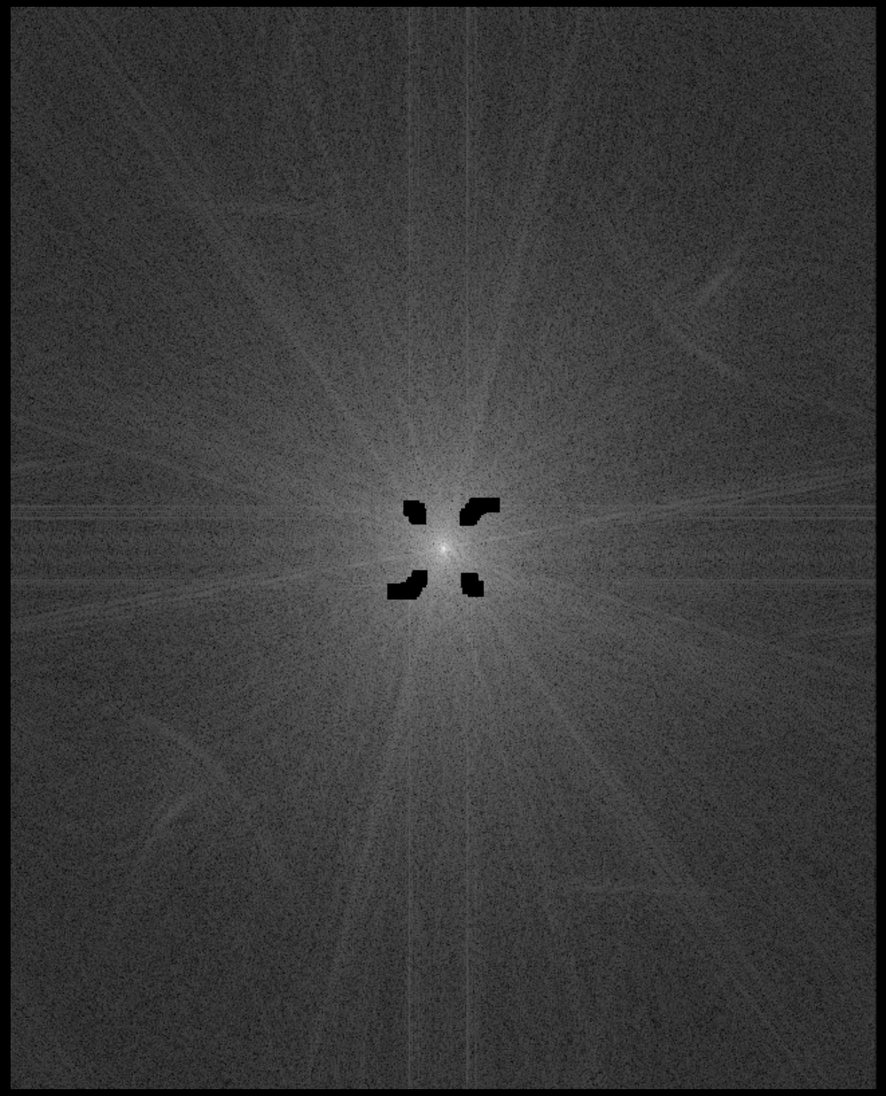


Frequency Spectrum Without and With Noise





Knocking out the noise - magic!



The Convolution Theorem

The Convolution Theorem

The convolution of two signals in the time (or spatial) domain is equivalent to the multiplication of their Fourier transforms in the frequency domain.

If you have two signals $f(t)$ and $g(t)$, their convolution in the time domain is represented as:

$(f * g)(t)$

The Fourier transforms of $f(t)$ and $g(t)$ are $F(\omega)$ and $G(\omega)$

According to the convolution theorem:

$$\mathscr{F}\{f(t) * g(t)\} = F(\omega) \cdot G(\omega)$$

The inverse is also the case.

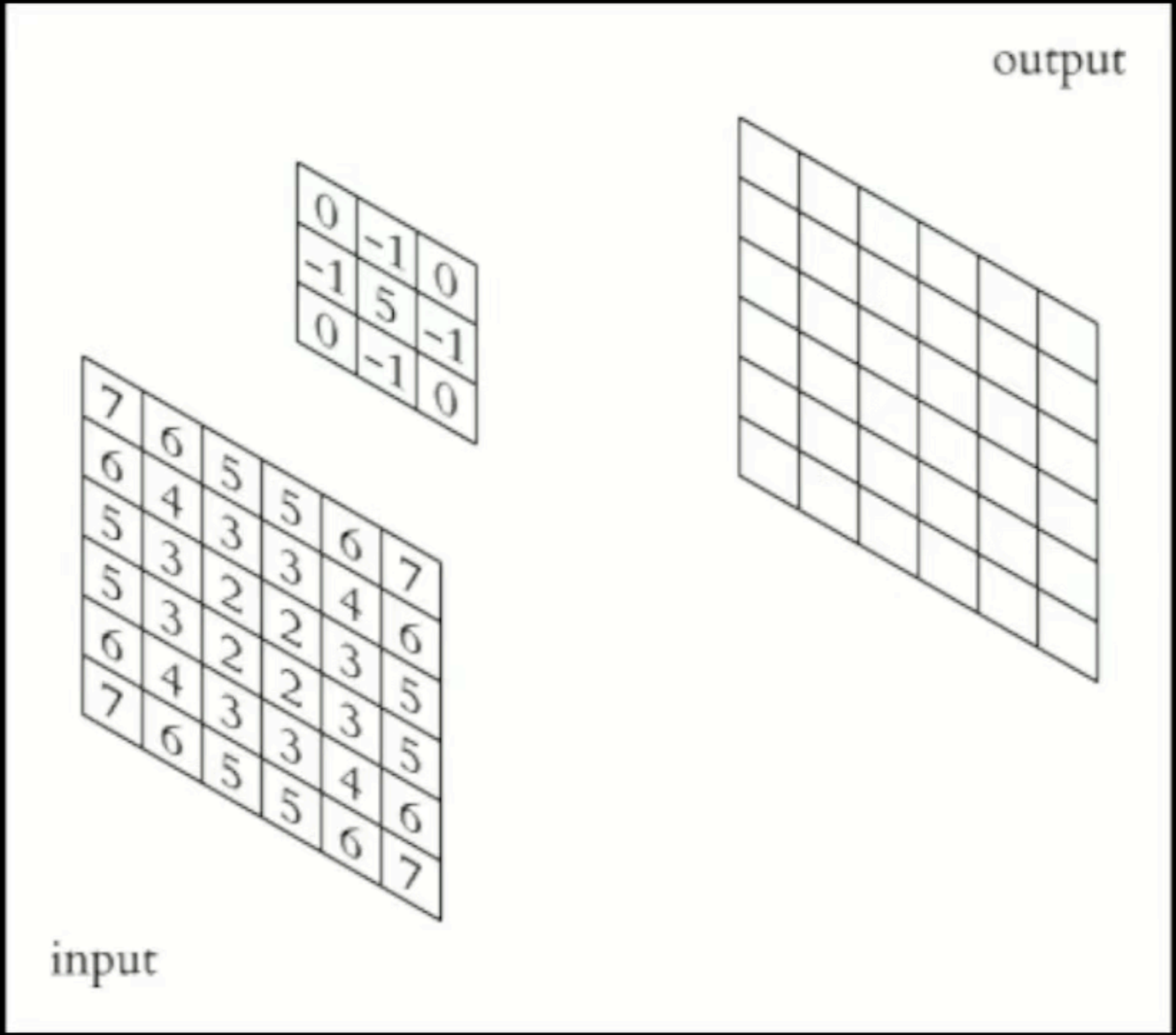
Where:

$\mathscr{F}$ represents the Fourier transform operation.

$*$ represents convolution in the time or spatial domain.

$\cdot$ represents multiplication in the frequency domain.

Convolution

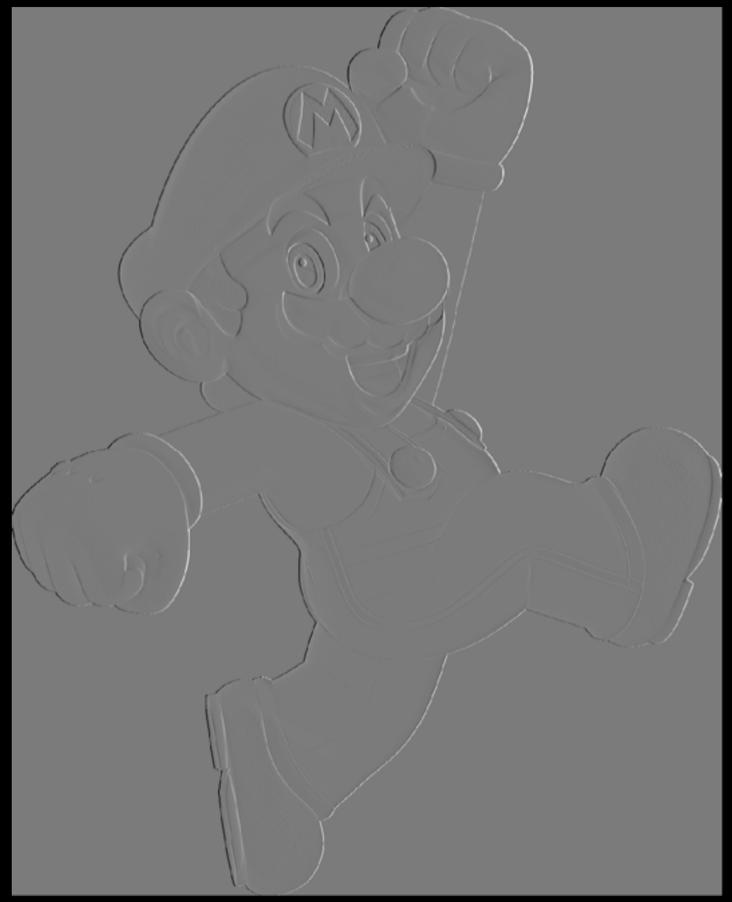


Michael Plotke, CC BY-SA 3.0 <https://creativecommons.org/licenses/by-sa/3.0>, via Wikimedia Commons

The Convolution Theorem

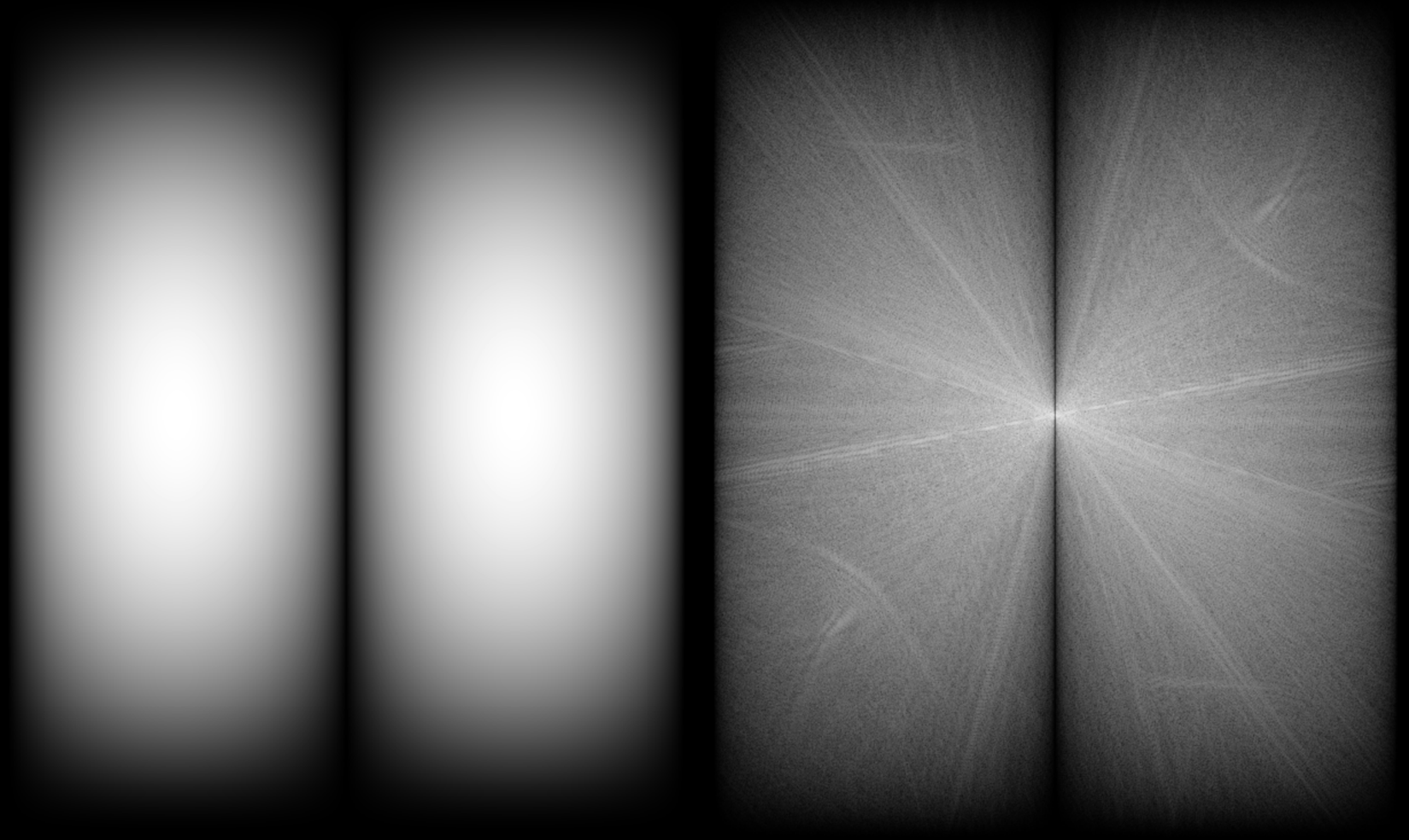
An example! Sobel Edge Detection

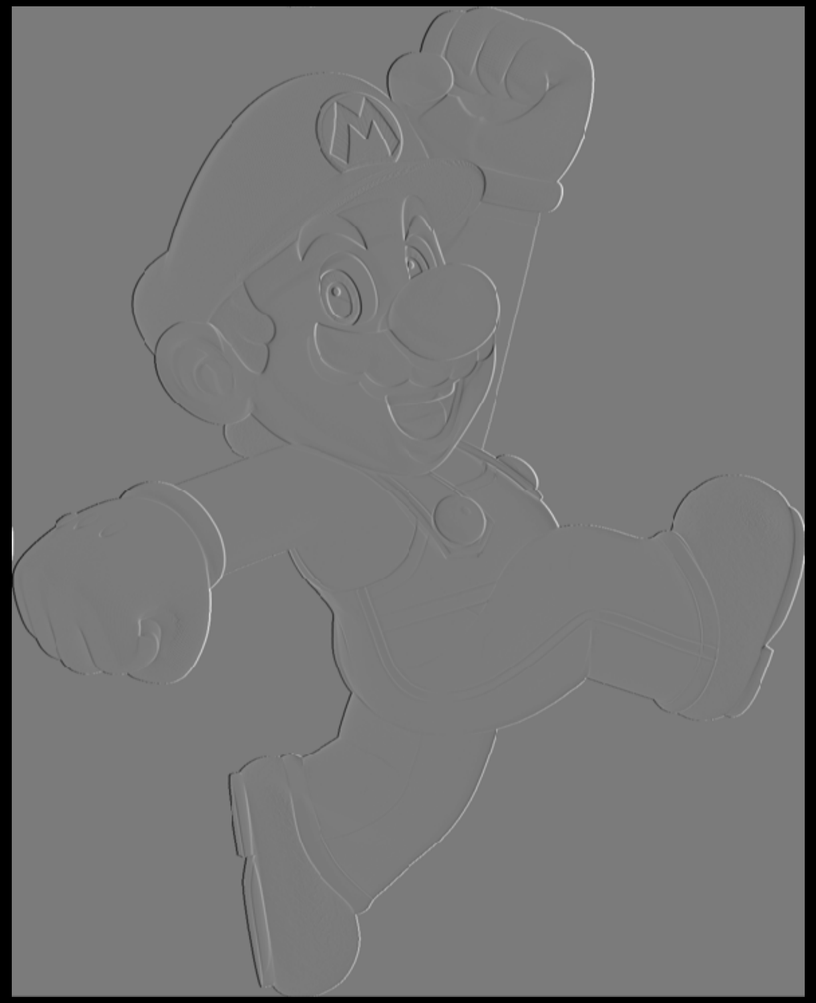
-1	0	1
-2	0	2
-1	0	1



The Convolution Theorem

An example! Sobel Edge Detection





The Convolution Theorem

But Why?

2D Convolution

An image is of size $N \times N$
and a kernel of size $M \times M$

Complexity is

$$O(N^2 \cdot M^2)$$

Convolution using FFT

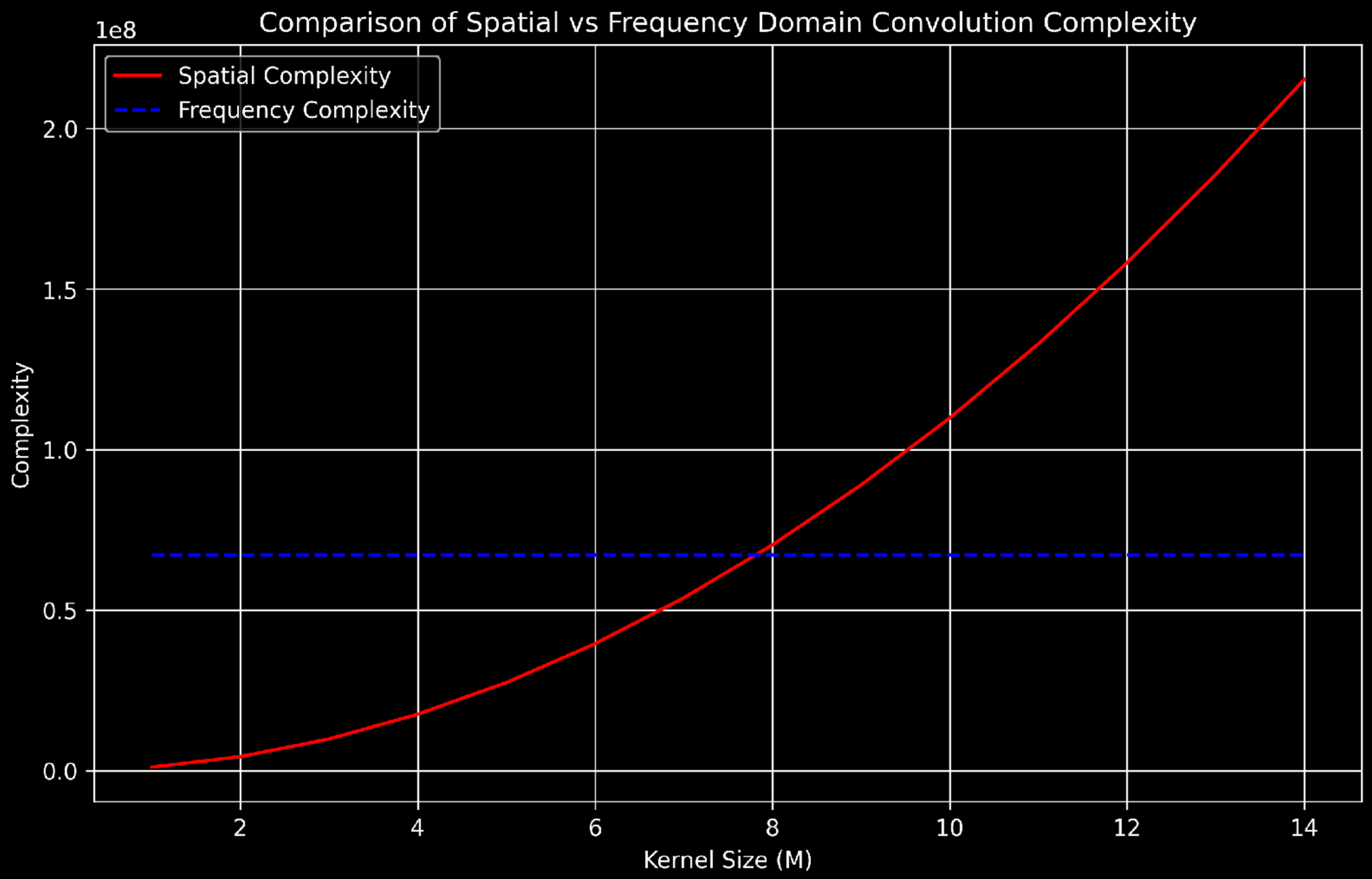
FFT of image - $O(N^2 log(N^2))$
FFT of kernel - $O(N^2 log(N^2))$

Multiplication - $O(N^2)$

IFFT of result - $O(N^2 log(N^2))$

Complexity is:

$$3 \times O(N^2 log(N^2)) + O(N^2)$$


Comparison of Spatial vs Frequency Domain Convolution Complexity
1e8
Spatial Complexity
Frequency Complexity
2.0
1.5
1.0
0.5
0.0
Complexity
2
4
6
8
10
12
14
Kernel Size (M)

The Identity Kernel

$$\begin{matrix} 0 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{matrix}$$

Sobel (Edge Detection)

$$\begin{matrix} -1 & 0 & 1 \\ -2 & 0 & 2 \\ -1 & 0 & 1 \end{matrix}$$

$$\begin{matrix} -1 & -2 & -1 \\ 0 & 0 & 0 \\ 1 & 2 & 1 \end{matrix}$$

Sobel (Edge Detection)

$$\begin{matrix} -1 & 0 & 1 \\ -2 & 0 & 2 \\ -1 & 0 & 1 \end{matrix}$$

$$\begin{matrix} -1 & -2 & -1 \\ 0 & 0 & 0 \\ 1 & 2 & 1 \end{matrix}$$

Gaussian Blur

1	4	6	4	1
4	16	24	16	4
6	24	36	24	6
4	16	24	16	4
1	4	6	4	1

Sharpen

$$\begin{matrix} 0 & -1 & 0 \\ -1 & 5 & -1 \\ 0 & -1 & 0 \end{matrix}$$

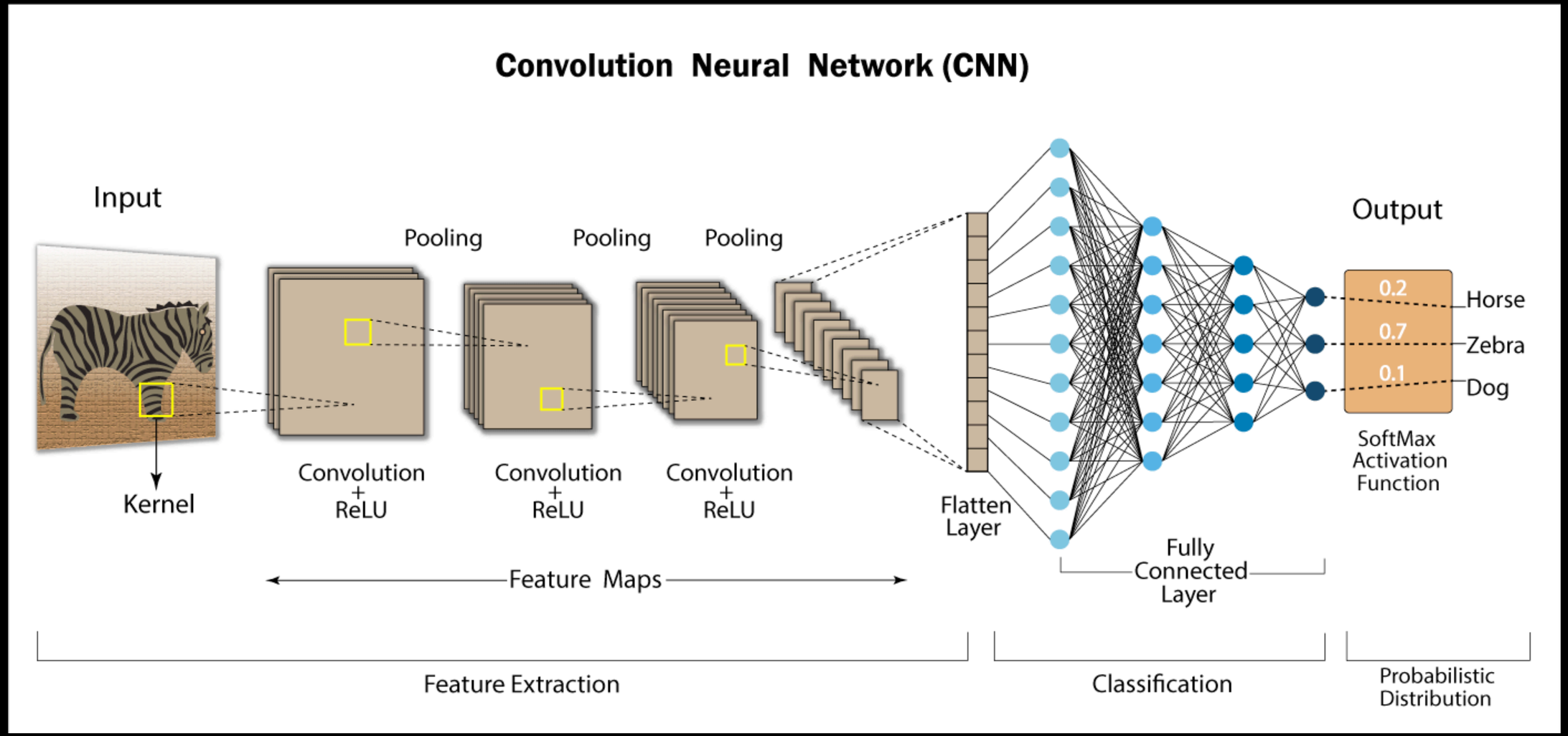
Feature Detection with Kernels

- Convolution kernels highlight specific features in images.
- Different kernels detect different types of features: edges, textures, corners, etc.

Convolutional Neural Networks

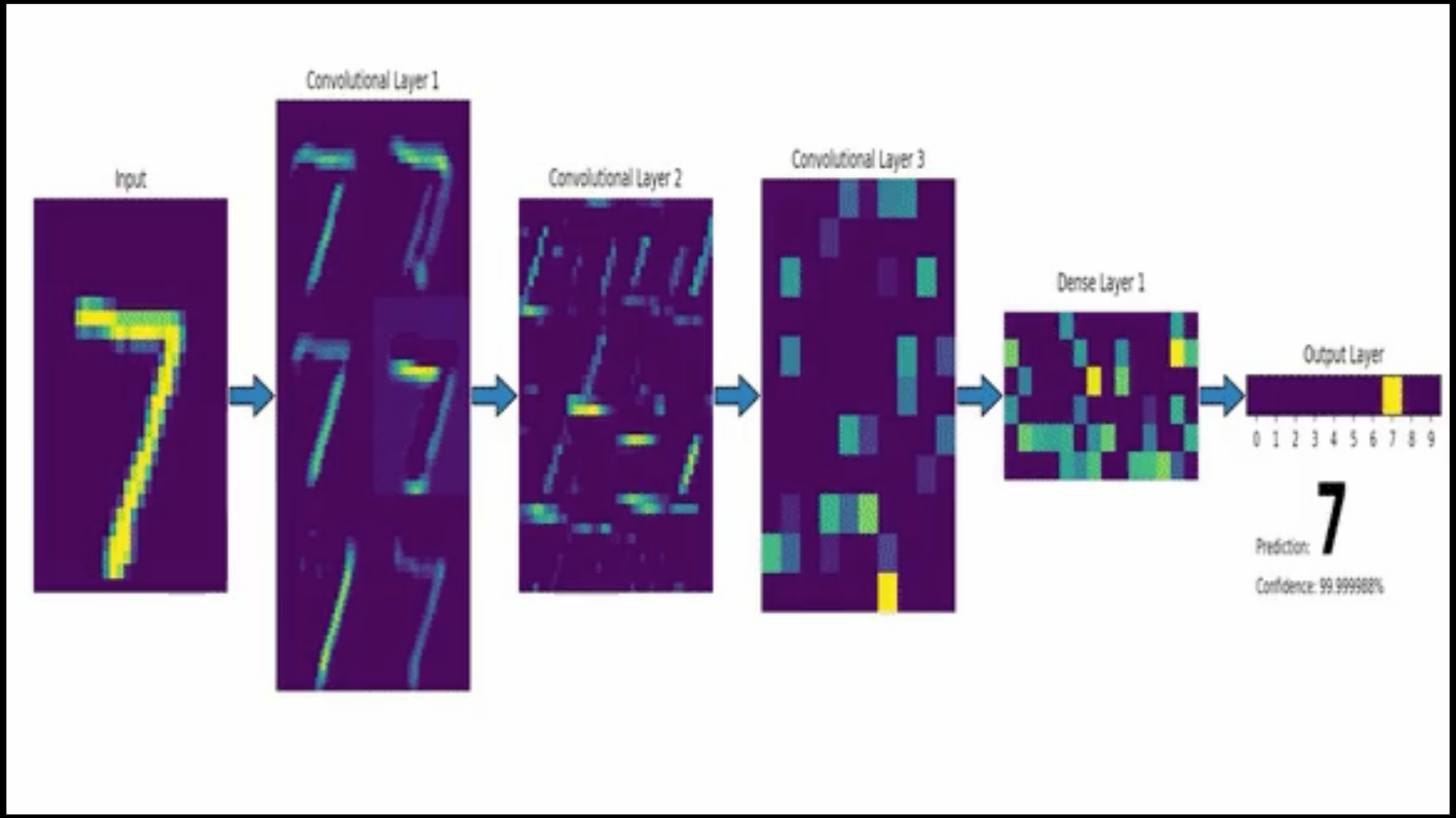
Convolutional Neural Networks extend this concept by automatically learning the best kernels for a given task.

Convolutional Neural Networks



https://blog.devgenius.io/convolutional-neural-networks-98e2067072a9

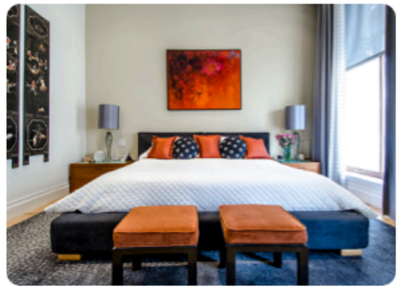
Convolutional Neural Networks



https://blog.devgenius.io/convolutional-neural-networks-98e2067072a9

Very nice - but I want to run it on my phone!

https://developer.apple.com/machine-learning/models/



FCRN-DepthPrediction
Depth Estimation

Predict the depth from a single image.

MNIST
Drawing Classification

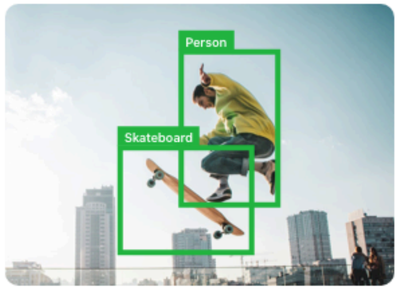
Classify a single handwritten digit (supports digits 0-9).

UpdatableDrawingClassifier
Drawing Classification

Drawing classifier that learns to recognize new drawings based on a K-Nearest Neighbors model (KNN).

DeeplabV3
Image Segmentation

Segment the pixels of a camera frame or image into a predefined set of classes.



YOLOv3
Object Detection

Locate and classify 80 different types of objects present in a camera frame or image.

MobileNetV2
Image Classification

The MobileNetv2 architecture trained to classify the dominant object in a camera frame or image.

Resnet50
Image Classification

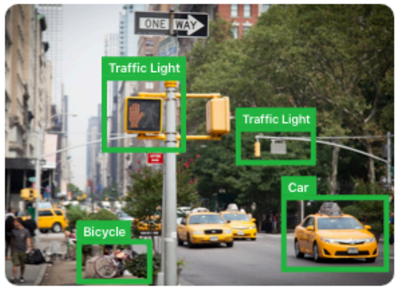
A Residual Neural Network that will classify the dominant object in a camera frame or image.

SqueezeNet
Image Classification

A small Deep Neural Network architecture that classifies the dominant object in a camera frame or image.

PoseNet
Pose Estimation

Estimates up to 17 joint positions for each person in an image.



YOLOv3-Tiny
Object Detection

Locate and classify 80 different types of objects present in a camera frame or image.

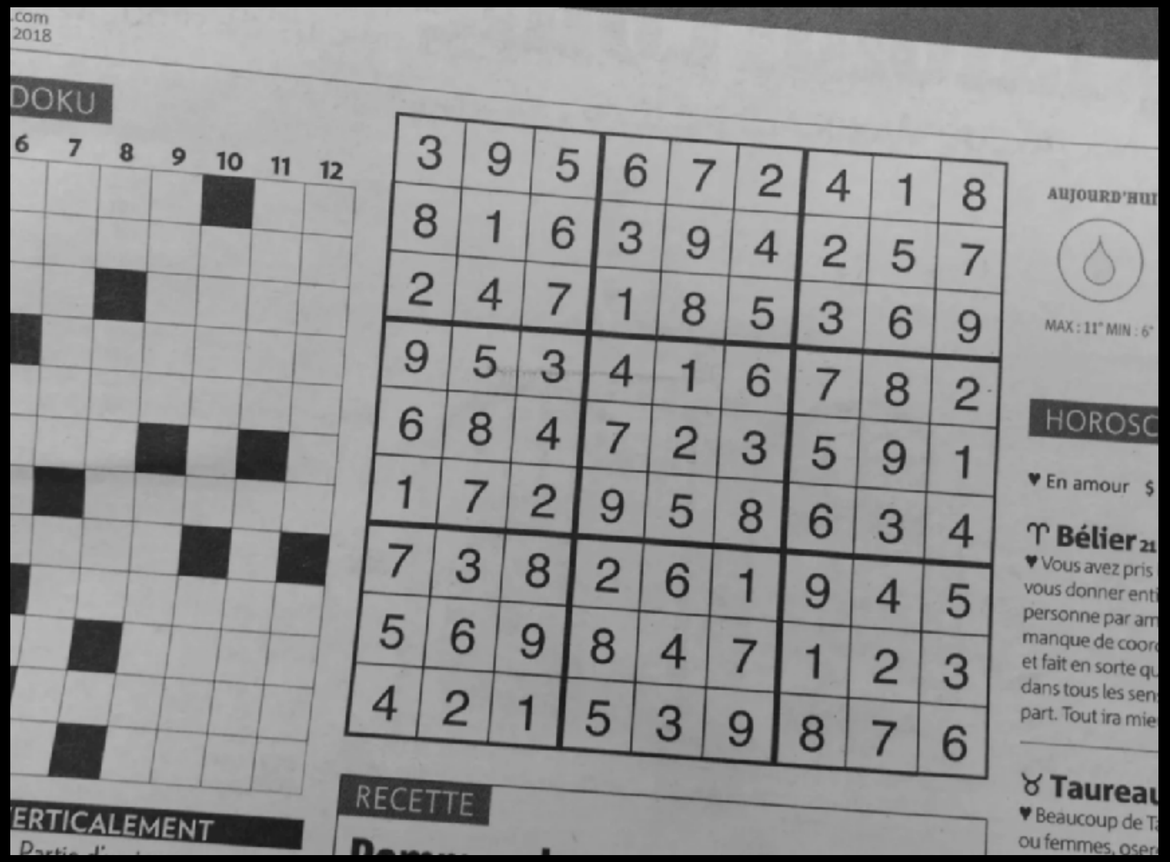
Image Segmentation

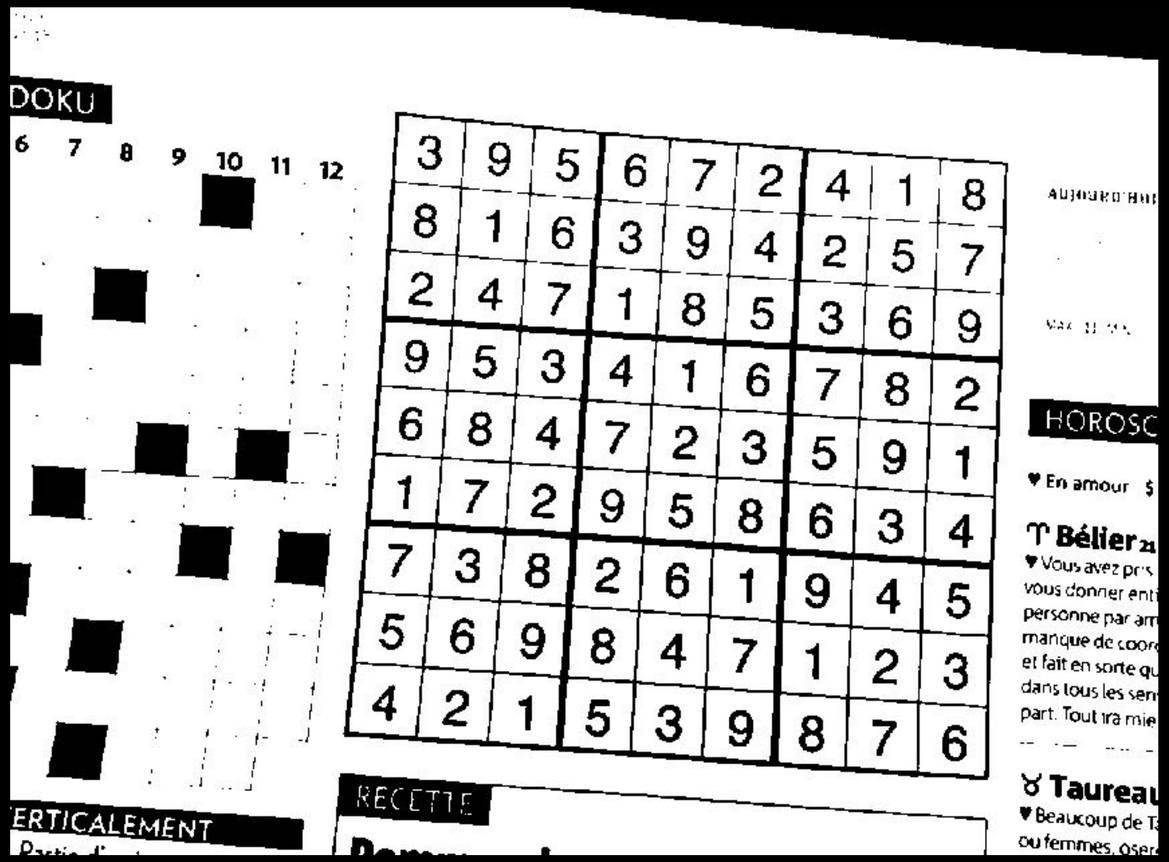
Dividing an image into multiple segments (or sets of pixels) to make the image simpler and more meaningful, transforming it into something that's easier to analyze.

Thresholding

```
value > X ? foreground : background
```

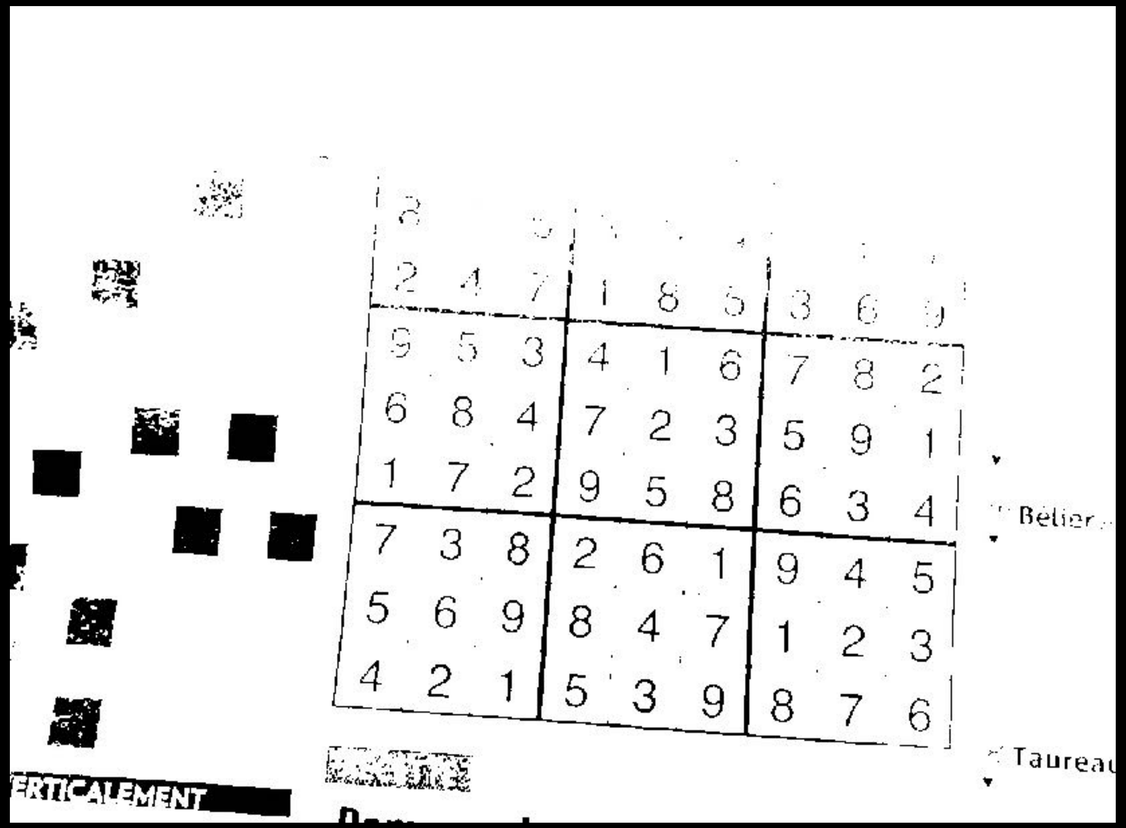
Binarizing an image based on a threshold value, separating it into foreground and background pixels.

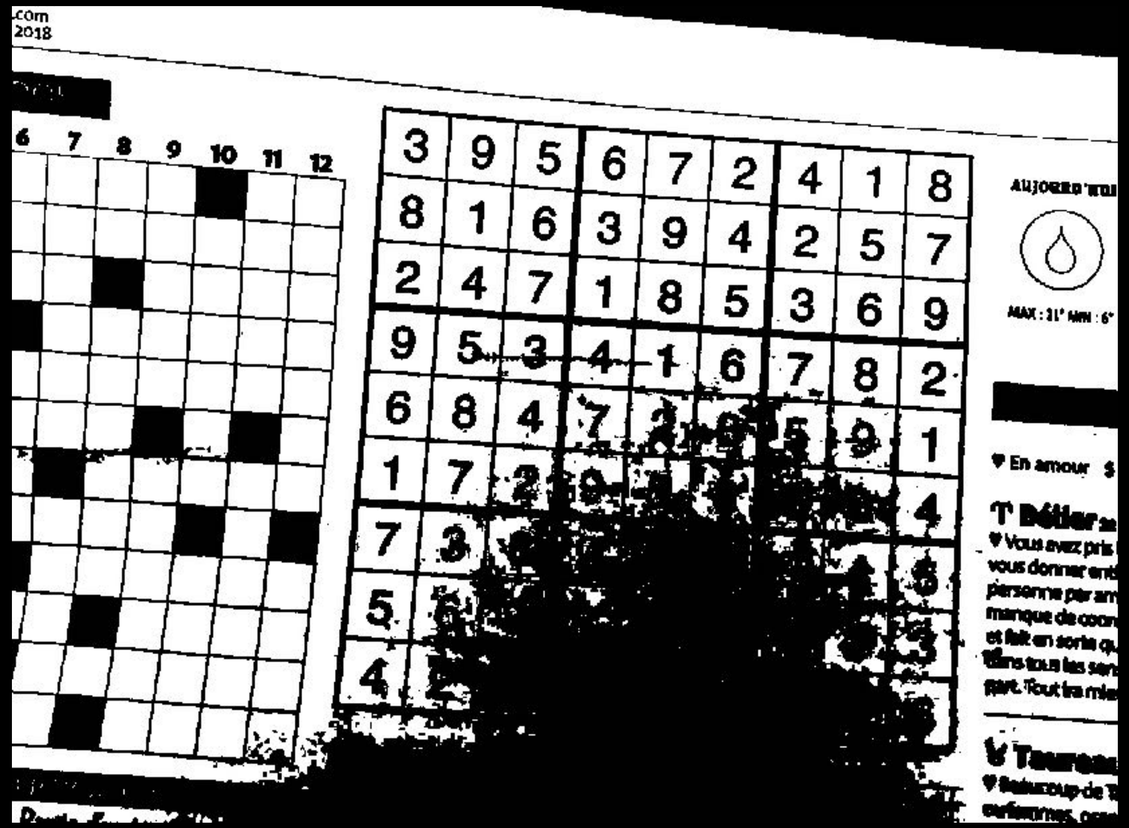




Global Threshold

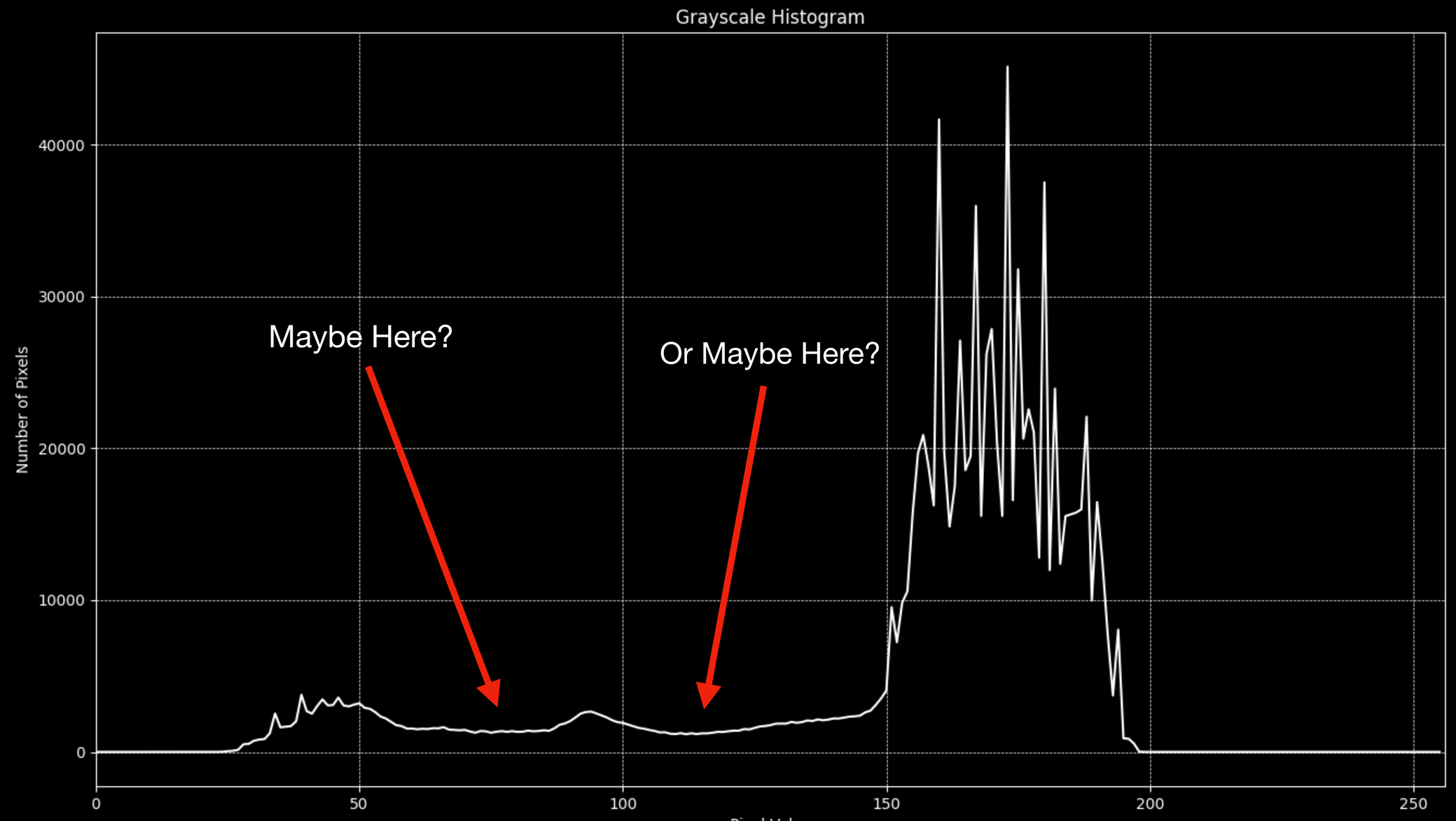
But what value to use for the threshold?





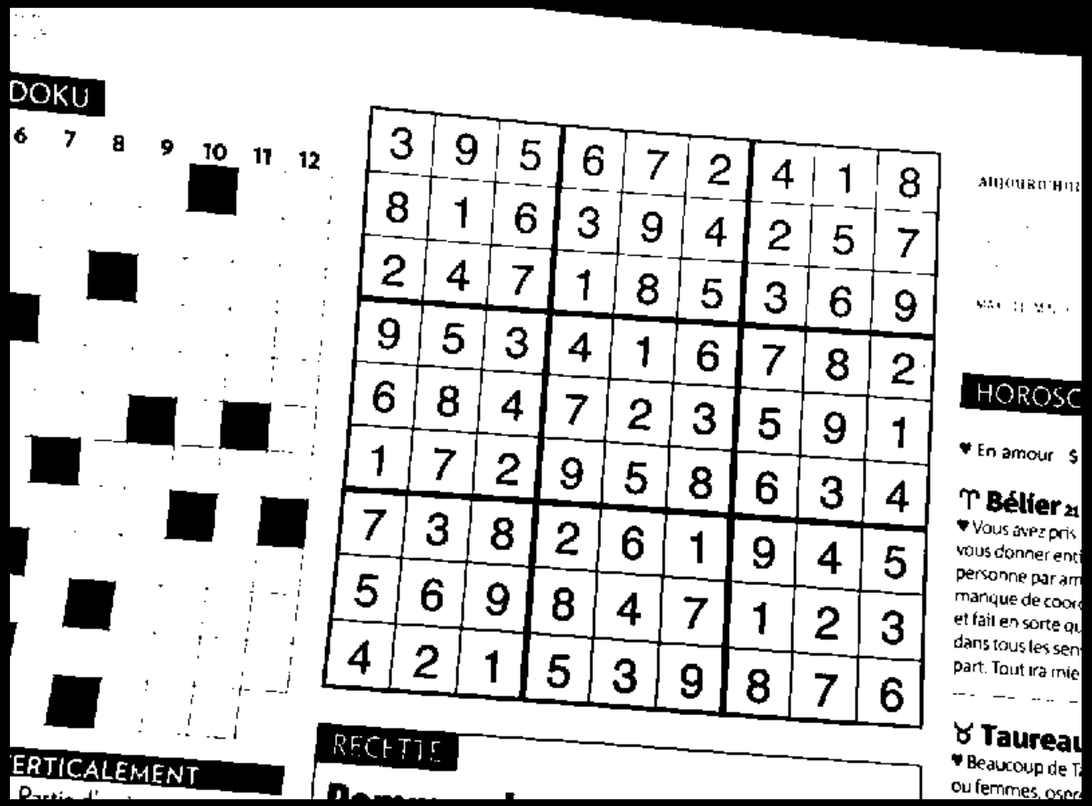
Histogram Can Give Some Clues

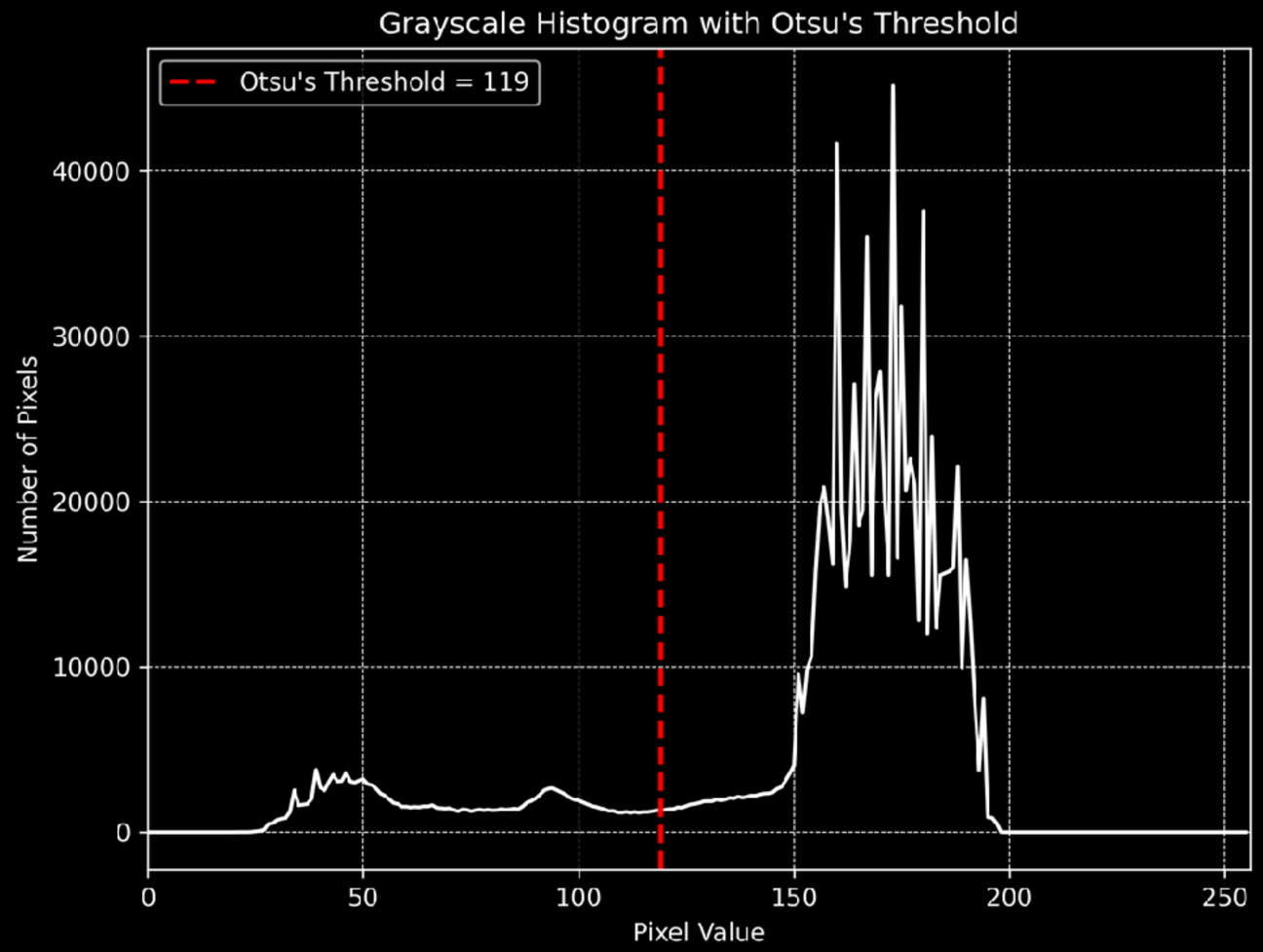
But this is a manual process and not always obvious



Can use Otsu’s Method

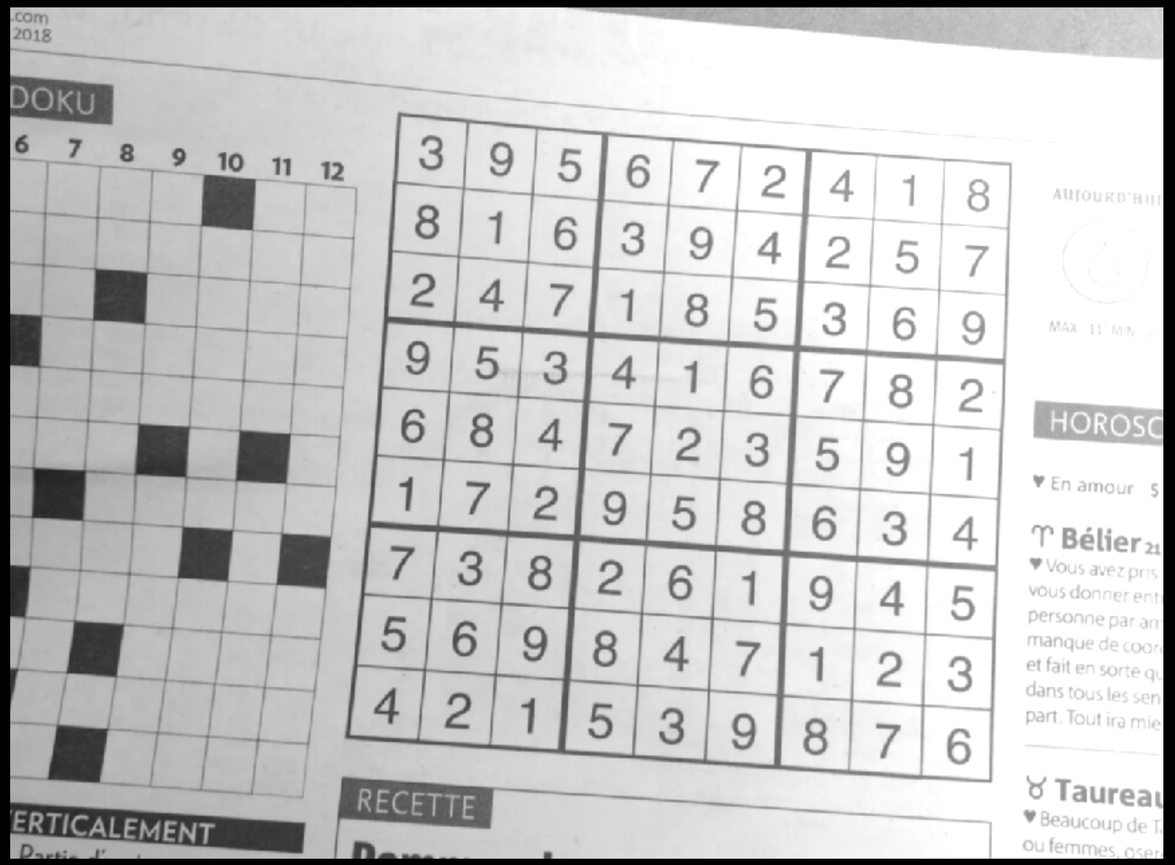
Built into OpenCV - so very easy to use!

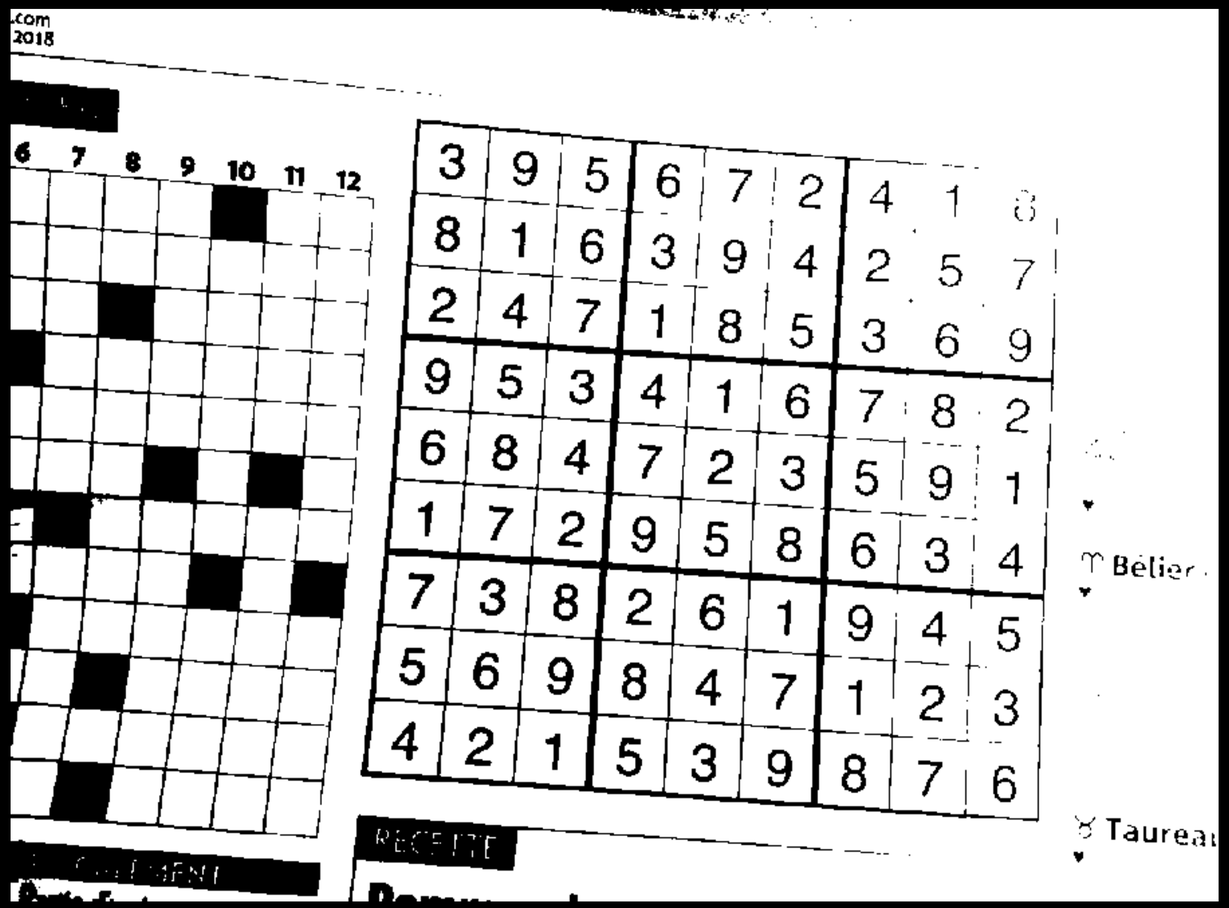




But Global Threshold Is Not Always Good

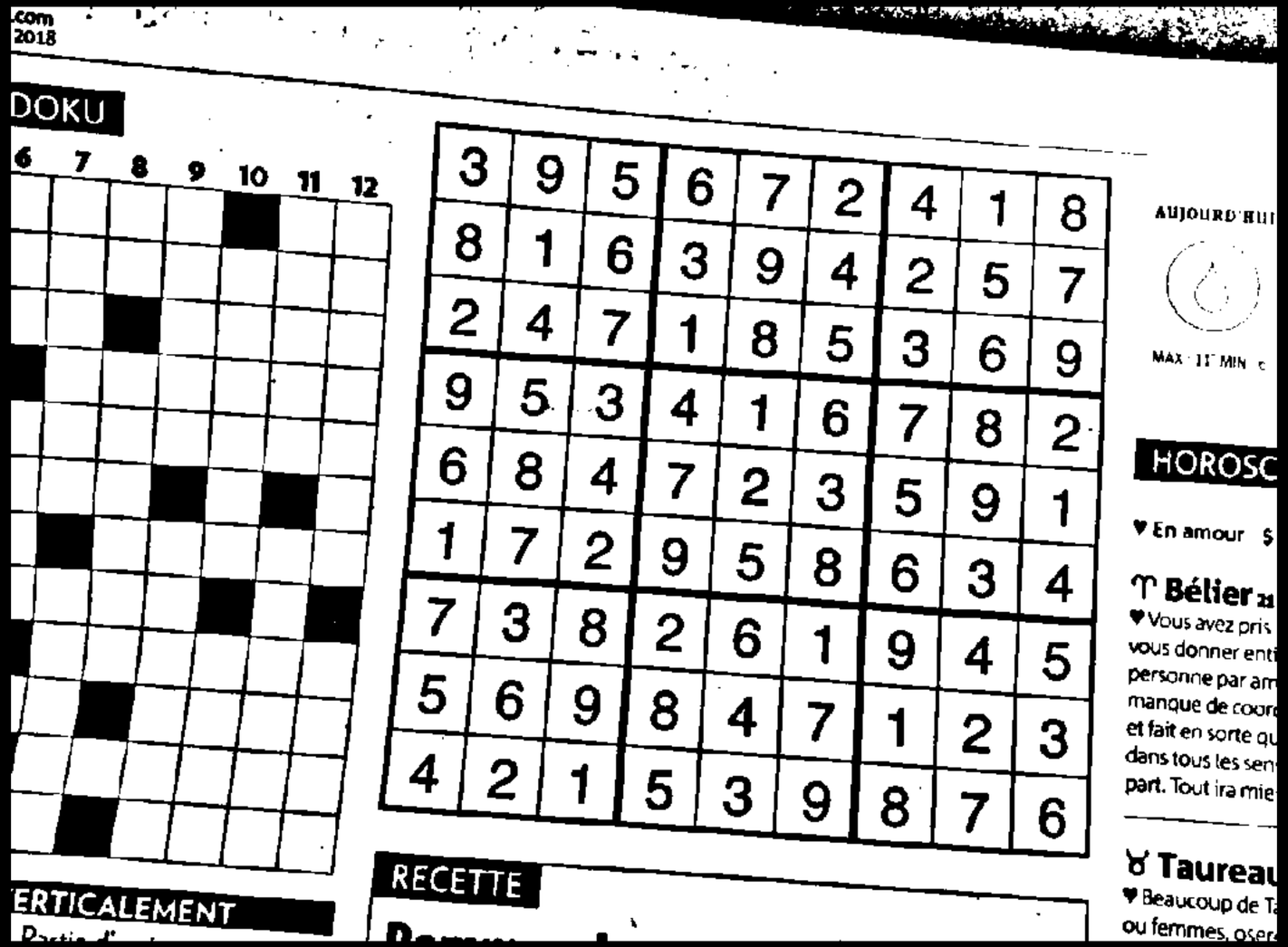
Doesn’t Handle Uneven Light and Shadow





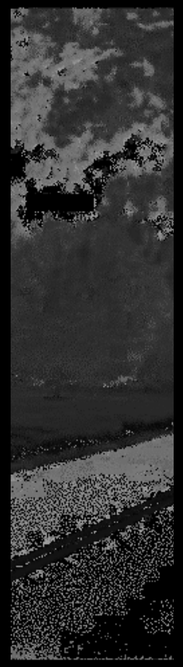
Adaptive Thresholding To The Rescue!

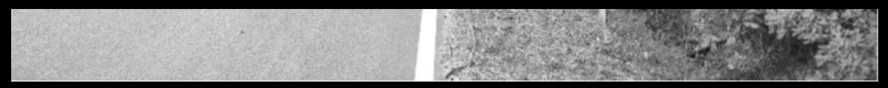
Compares Each Pixel With The Local Area



Segmenting Based On Colour

This is going to blow your mind!





Segmentation is a BIG subject!

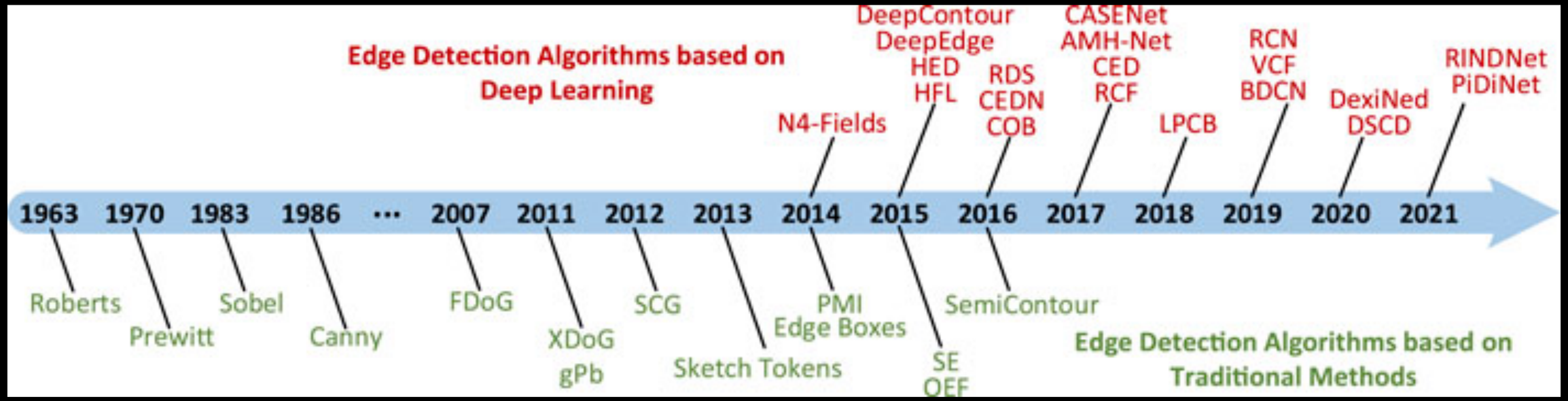
Further reading for the interested…

- **Region-Based Segmentation**
  - Divide an image into uniform regions based on similarity or homogeneity criteria.
- **Edge Detection-Based Segmentation**
  - Segment objects by identifying and connecting their boundaries.
- **Watershed Algorithm**
  - Separate touching objects by visualizing the image gradient as topographic basins.
- **Segmentation using Clustering**
  - Group pixels based on intensity or color similarities.

Edge Detection

Edge Detection

It’s an evolving field!



https://www.frontiersin.org/articles/10.3389/frsip.2022.826967/full

Edge Detection

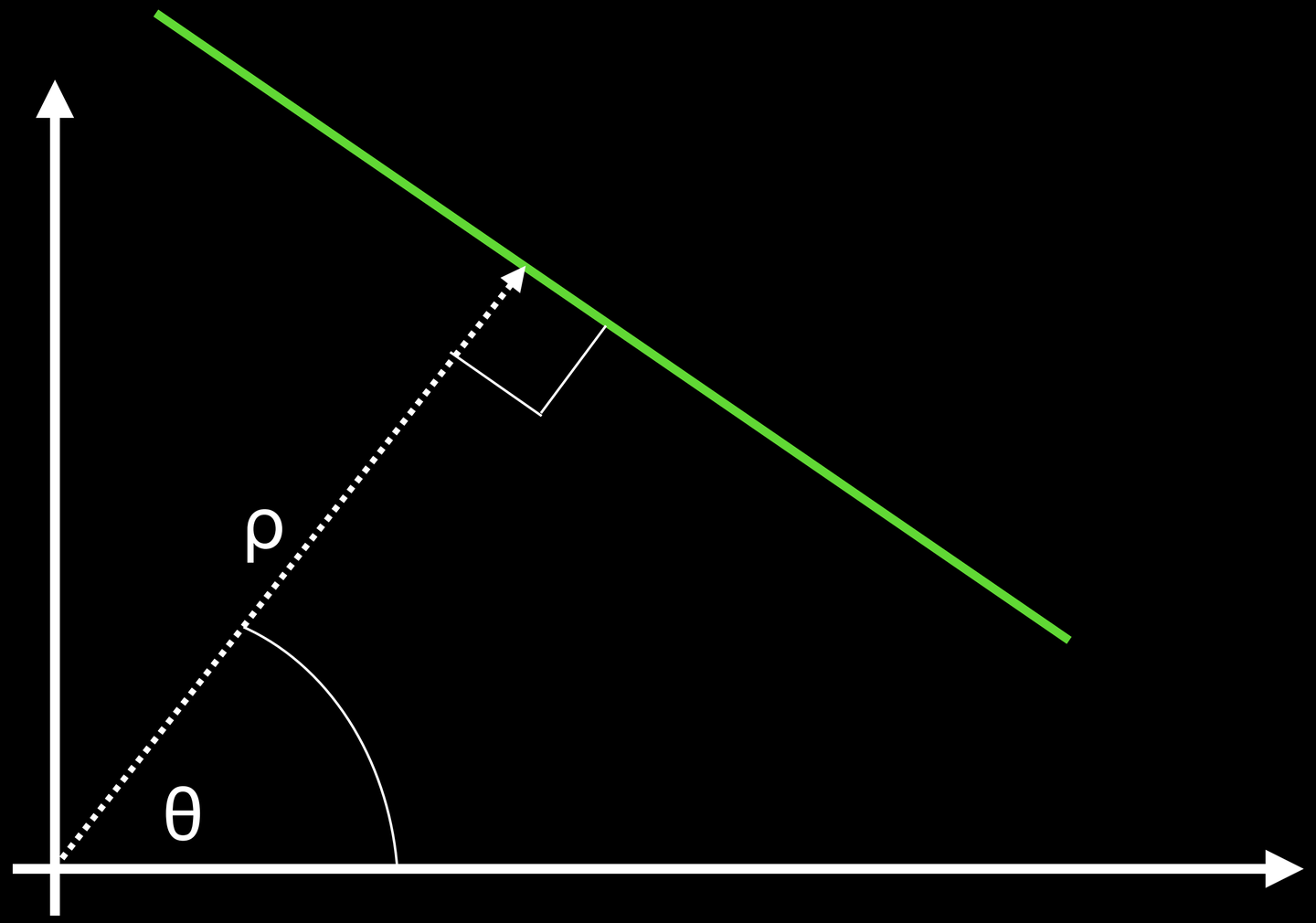
Sobel

Canny

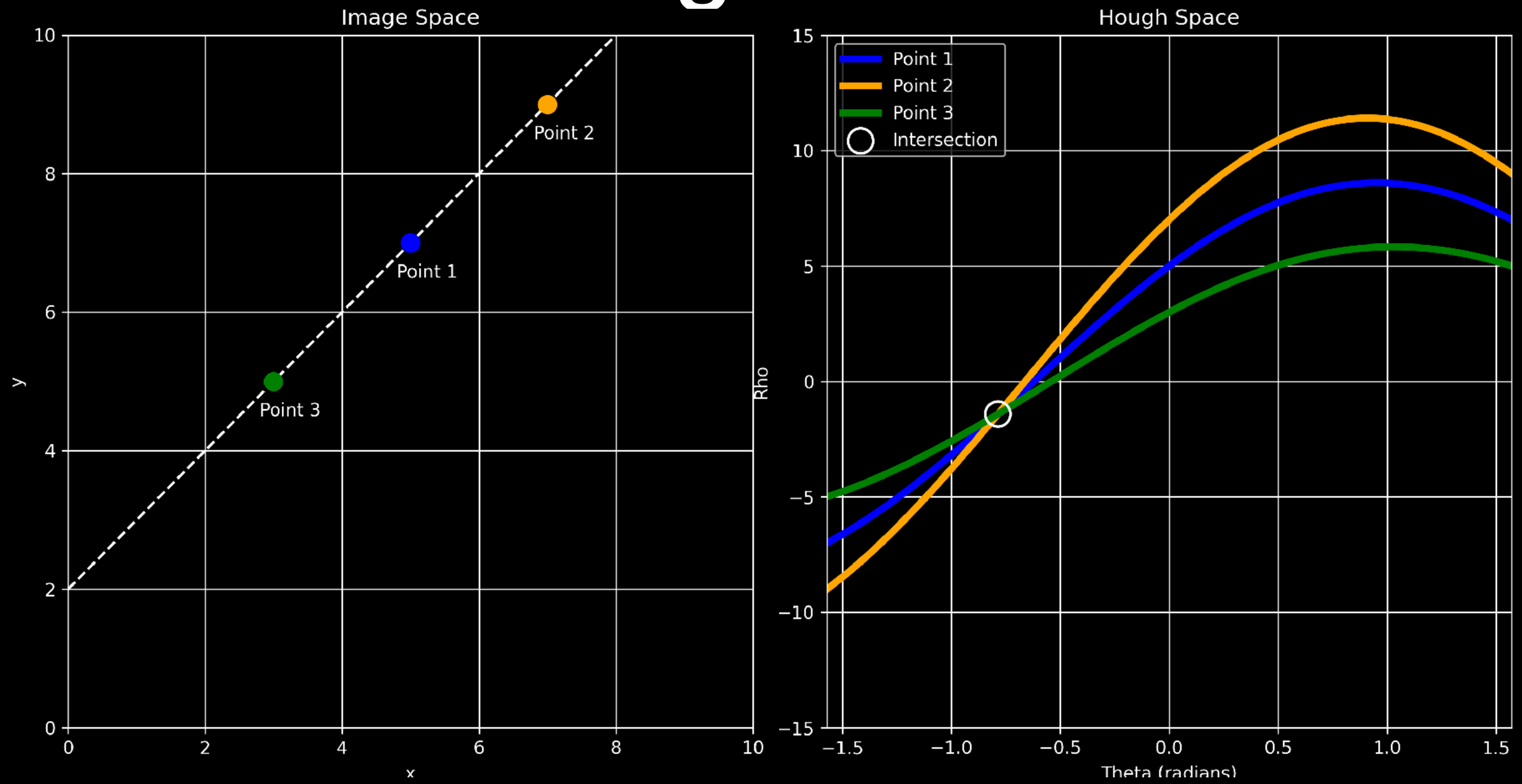
HED

The Hough Transform

The Hough Transform

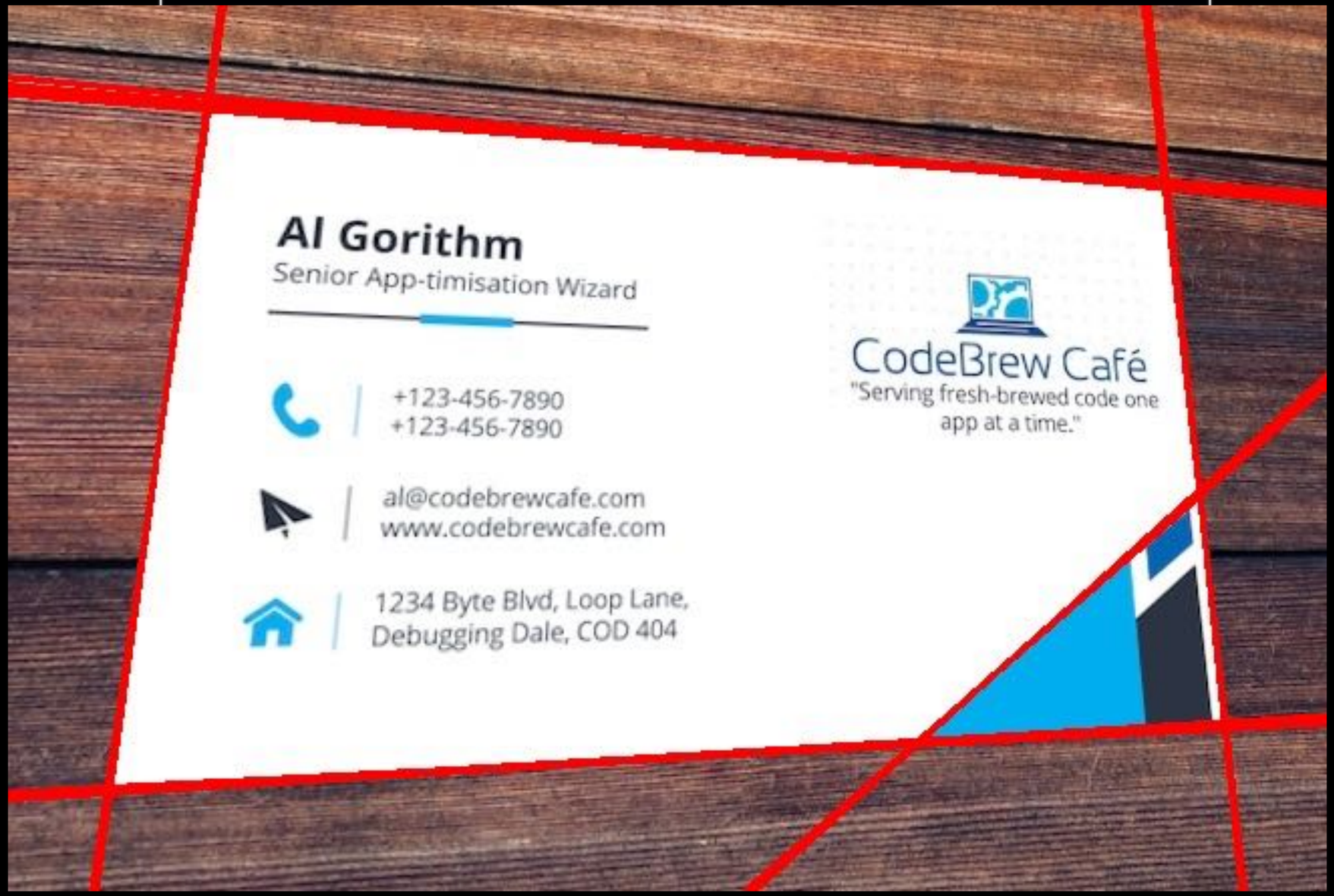


The Hough Transform



The Hough Transform

Hough Space

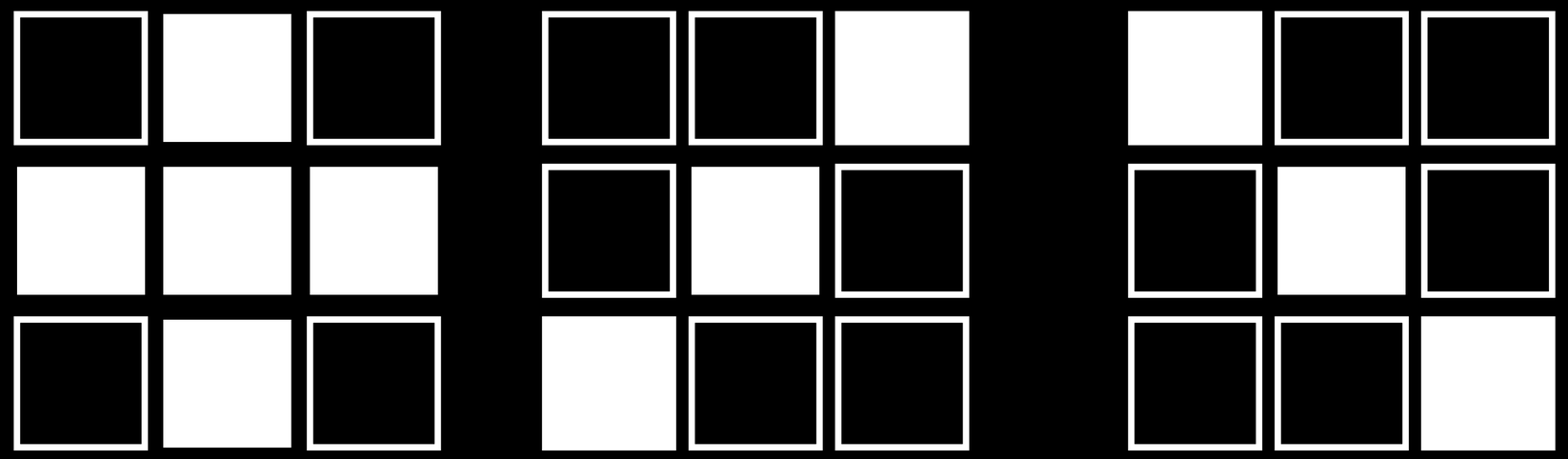


VNDetectRectangles Request

Morphological Operations

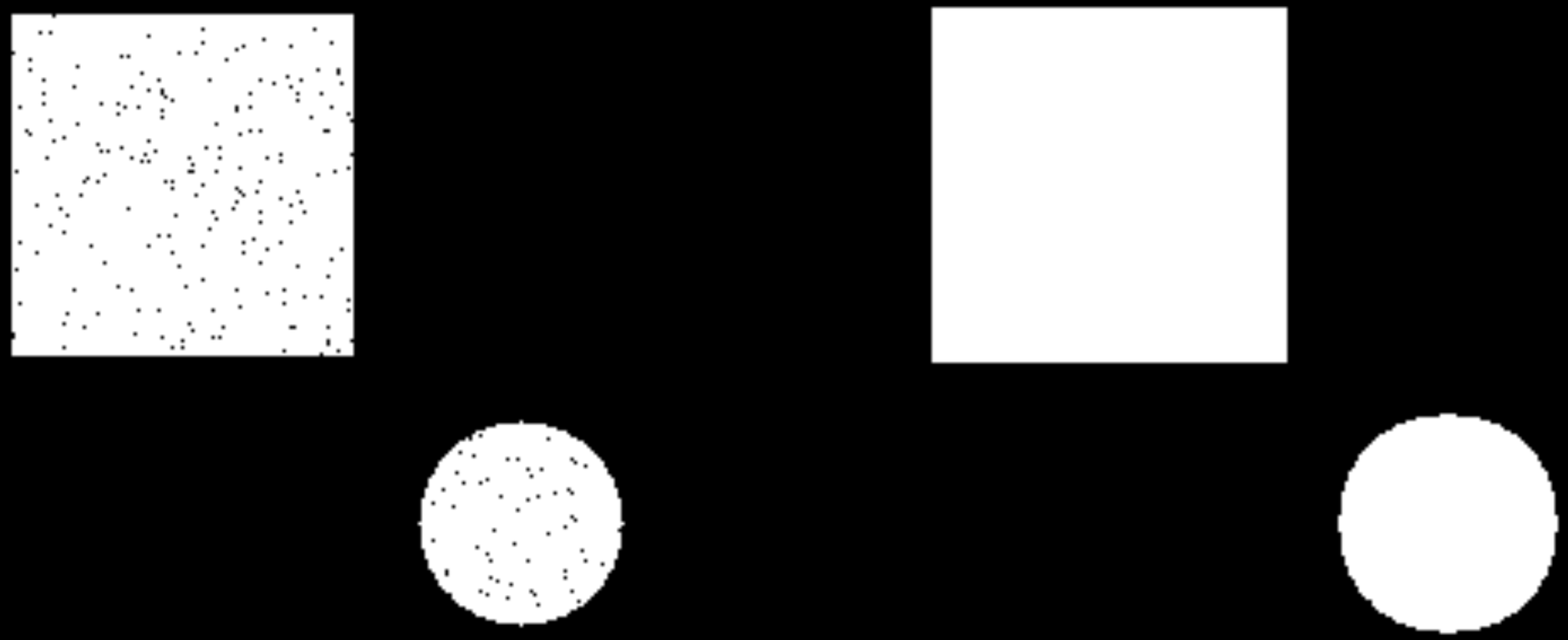
Techniques for processing image structures, primarily used on binary images.

- Rooted in set theory
- Uses a structuring element to probe and modify input image



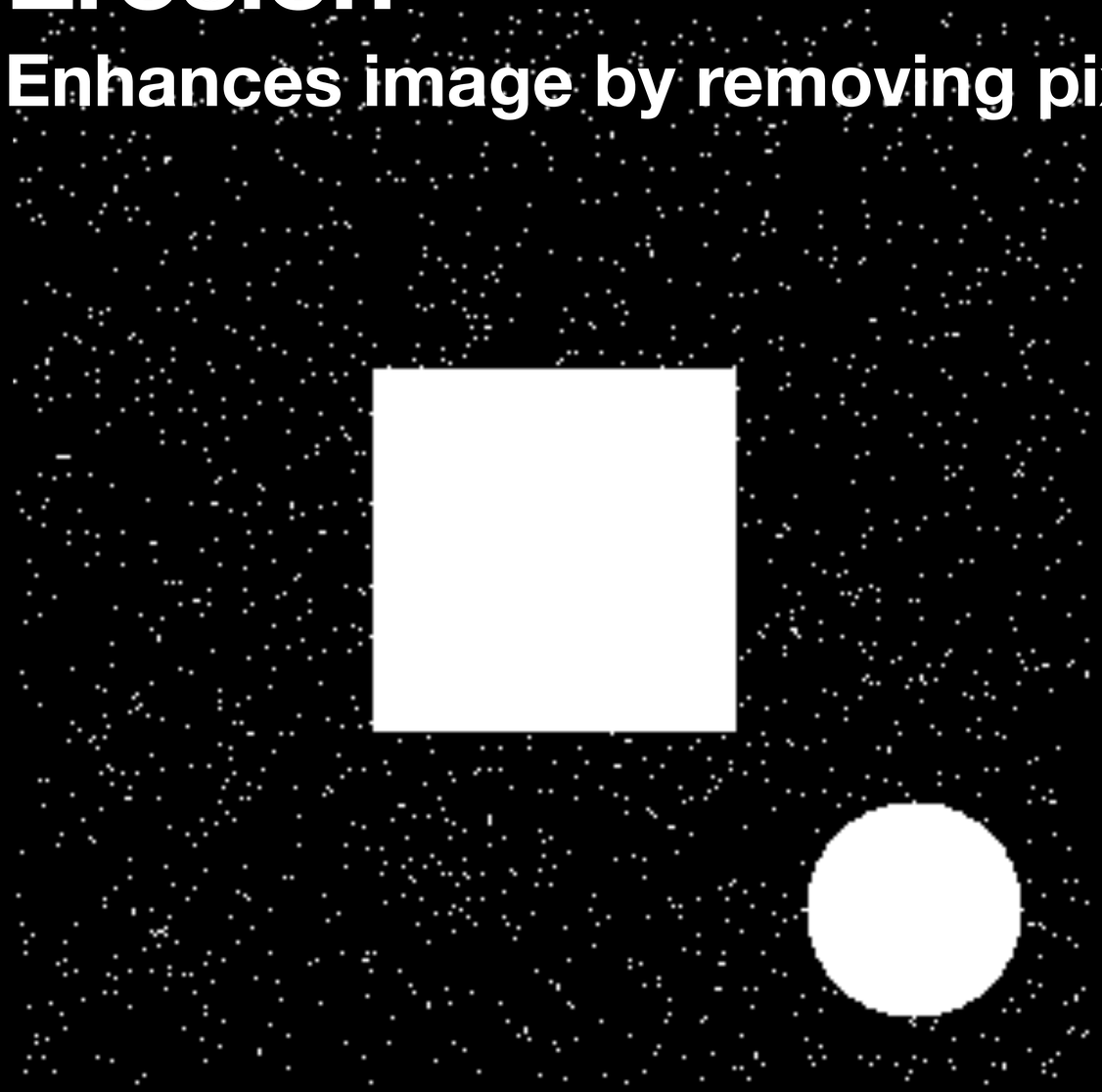
Dilation

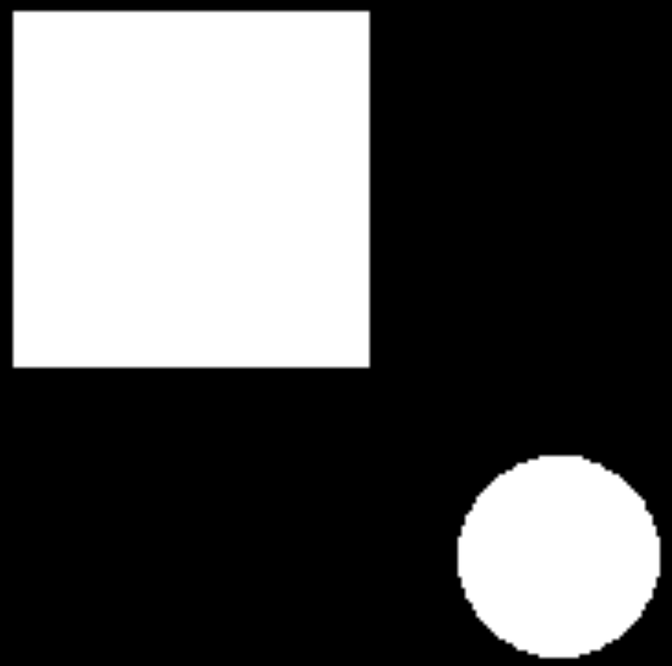
Enhances image by adding pixels



Erosion

Enhances image by removing pixels





Opening

Good for separating objects that should not be connected

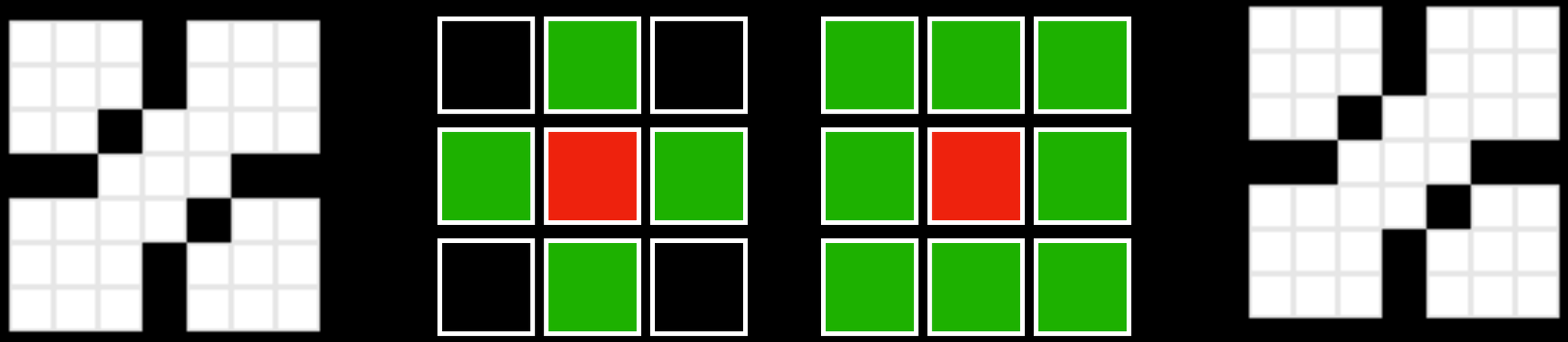
Closing

Good for connecting objects that are disconnected

Connected Component Analysis

Label connected regions in a binary image

- Starts with a pixel and then explores its neighbors (either 4-connectivity or 8-connectivity)
- Continues the exploration until all connected pixels are visited and labeled
- Moves to the next unlabeled pixel and repeats the process



By André Karwath aka Aka - Own work, CC BY-SA 2.5, https://commons.wikimedia.org/w/index.php?curid=481652

Connected Component Analysis

Label connected regions in a binary image

- Counting objects in a scene.
- Pre-processing step for object recognition.
- Cleaning up segmentation results (removing small blobs).
- Analyzing spatial relationships between objects.

Transformations

Scaling

Make big, make small

Rotation

You spin me round…

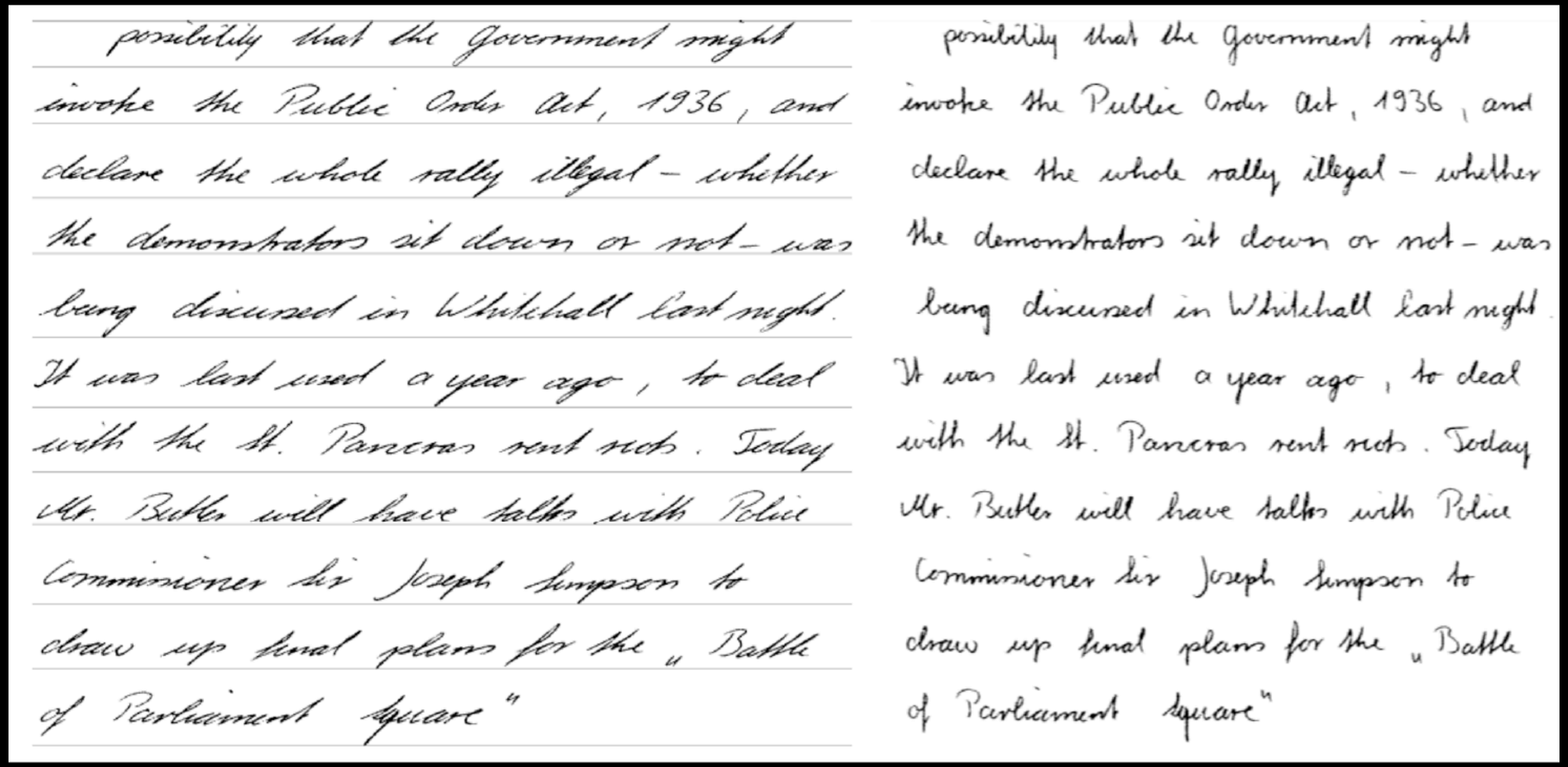
Shear

Taking a slant on reality

Shear

Applications

Kavallieratou, E.; Likforman-Sulem, L.; Vasilopoulos, N. Slant Removal Technique for Historical Document Images. J. Imaging 2018, 4, 80. https://doi.org/10.3390/jimaging4060080



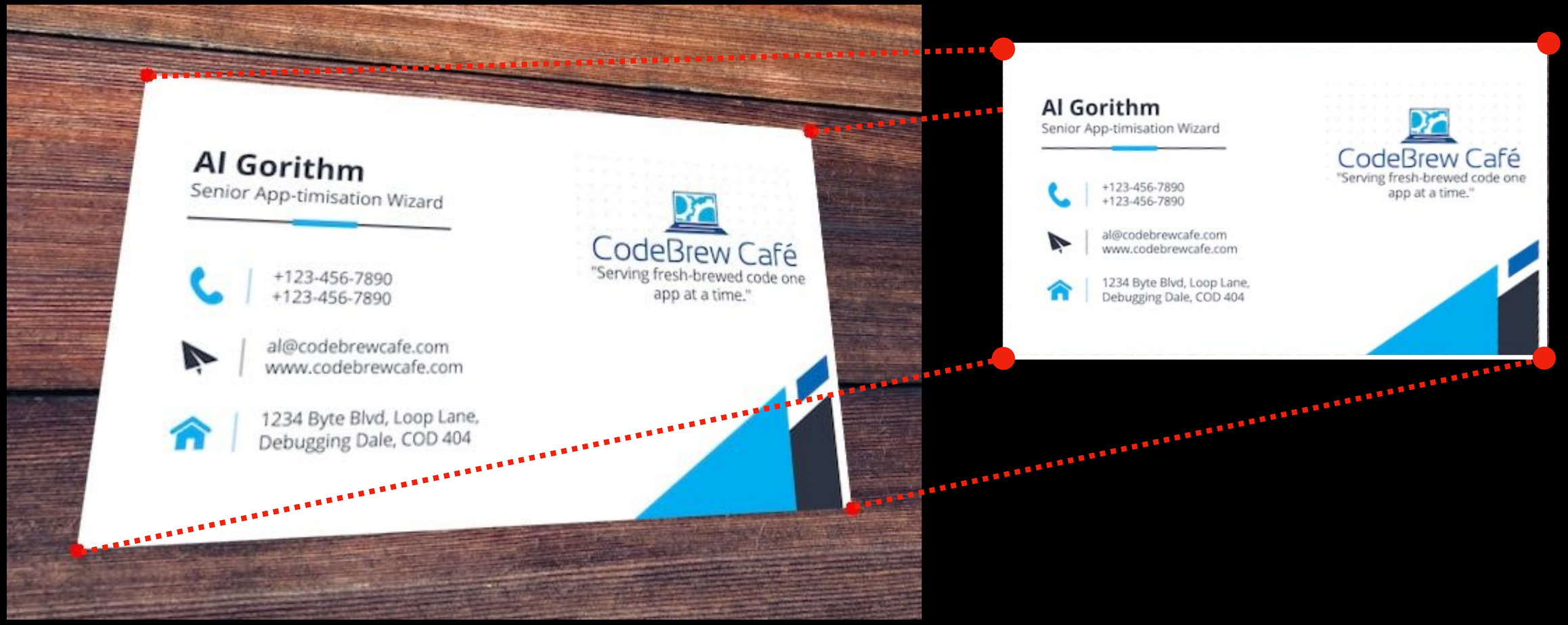
Homographic Transformation

Normally referred to as Homography

- **Perspective Correction**: Adjusting images taken at an angle to appear as if they were taken from the front. For example, making a photo of a painting or a sign appear as if it was taken head-on, even if shot from the side.
- **Image Stitching**: Combining two or more images with overlapping fields of view to produce a wide-angle panorama.
- **Augmented Reality**: Overlaying virtual 3D objects onto a real-world 2D plane, ensuring the virtual objects respect the perspective of the scene.
- **Simulation of Planar Camera Motion**: Virtually changing the viewpoint of a scene, like in video stabilization or creating "dolly zoom" effects.
- **3D Plane Reconstruction**: Reconstructing a 3D planar surface from multiple images.

Computing a Homography

- **Need at least four points in both image planes**



Computing a Homography

$$\begin{bmatrix} x' \\ y' \\ 1 \end{bmatrix} = \mathbf{H} \begin{bmatrix} x \\ y \\ 1 \end{bmatrix} = \begin{bmatrix} h_{11} & h_{12} & h_{13} \\ h_{21} & h_{22} & h_{23} \\ h_{31} & h_{32} & h_{33} \end{bmatrix} \begin{bmatrix} x \\ y \\ 1 \end{bmatrix}$$

There’s so so much more…

I Love it - But How Do I Do It?

What do you get out of the box?

- **vImage** - high-performance image processing framework
  - Convolutions
  - Geometric transformations
  - Histogram operations
  - Morphological transformations
  - etc…
- **Core Image**
  - Built in filters
  - Custom filters
- **Vision**
  - Face detection
  - Pose detection
  - Text detection
  - Barcodes
  - Image Registration
  - Feature Tracking
  - Custom Core ML models
- **Create ML**
  - Design and train models

And If That’s Not Enough!

OpenCV - Open Source Computer Vision Library

- Image Processing
- Feature Detection and Description
- Object Detection
- Image Segmentation
- Machine Learning
- Video Analysis
- Camera Calibration and 3D Reconstruction
- Integration with Deep Learning Frameworks

Cross-platform - Windows, Linux, macOS, iOS, and Android, and it provides interfaces for C++, Python, Java, and MATLAB.

You still want more!?

Get the raw pixels and process them yourself!

```
var image = Image<RGBA<UInt8>>(named: "ImageName")!

let pixel: RGBA<UInt8> = image[x, y]
image[x, y] = RGBA(red: 255, green: 0, blue: 0, alpha: 127)
image[x, y] = RGBA(0xFF00007F)

// Iterates over all pixels
for pixel in image {
    // ...
}
```

https://github.com/nsuinteger-momatu/swift-image

https://developer.apple.com/documentation/avfoundation/avcapturevideodataoutput

How to tackle a project - Questions to Ask…

- What are you actually trying to do?
- Do you have sample images?
- How constrained is the problem?
- How robust does the solution need to be?
- What are the “realtime” requirements?
- On device/offline or processed by a server?

How to tackle a project - Questions for you…

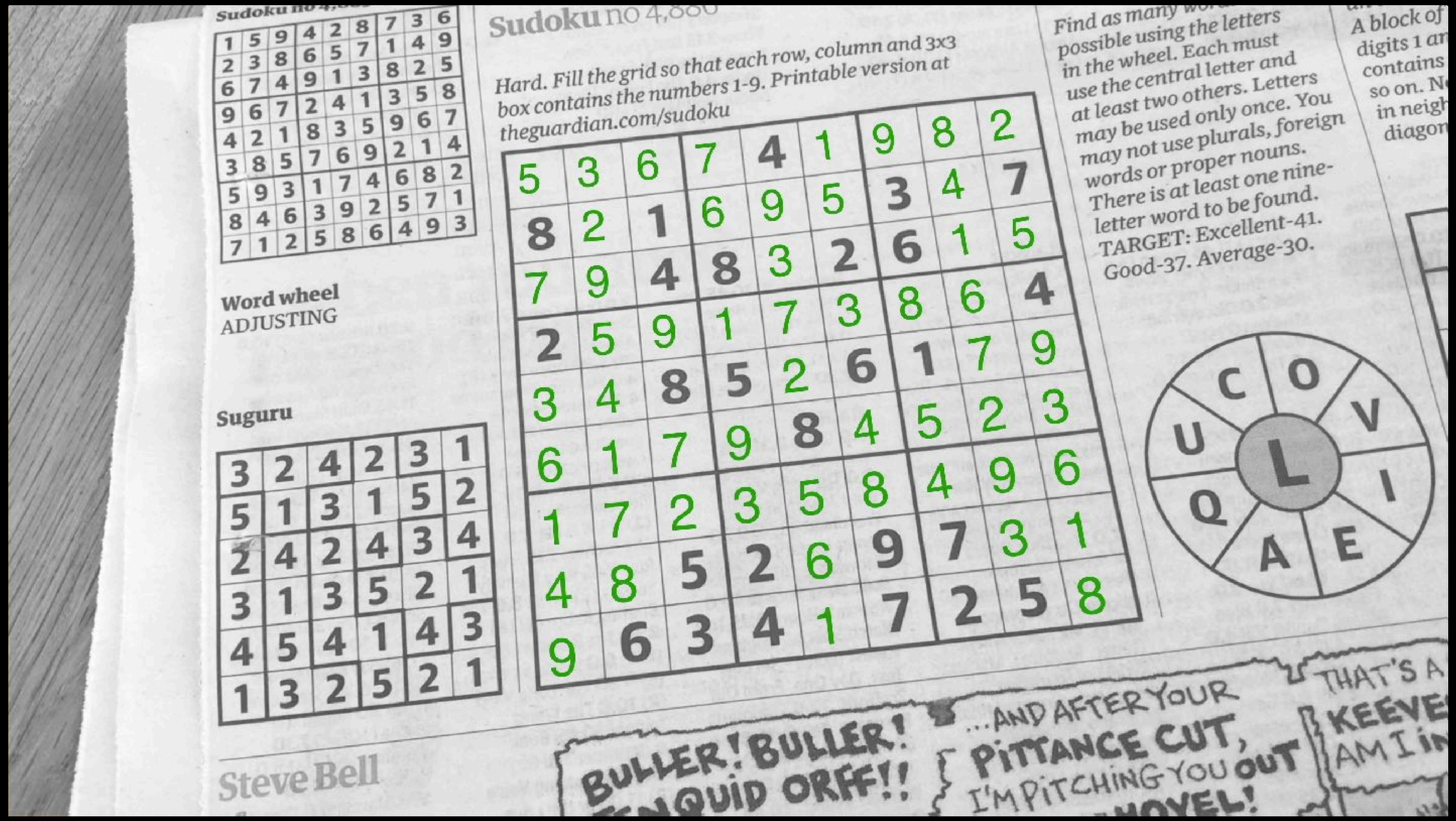
- Does it seem possible/feasible?
- Can you see the shape of the problem?
- Is it an already “solved” problem?
- Is it a deep learning problem?
- Is it a classic image processing problem?
- Or some combination of the two?
- Is it a research project?

General Approach

- Get some sample image
- Inspect the images
- Is there any preprocessing required?
- Try some out of the box algorithms
- Search for similar problems
- Create a rough end to end processing pipeline
- Prototype on your target system

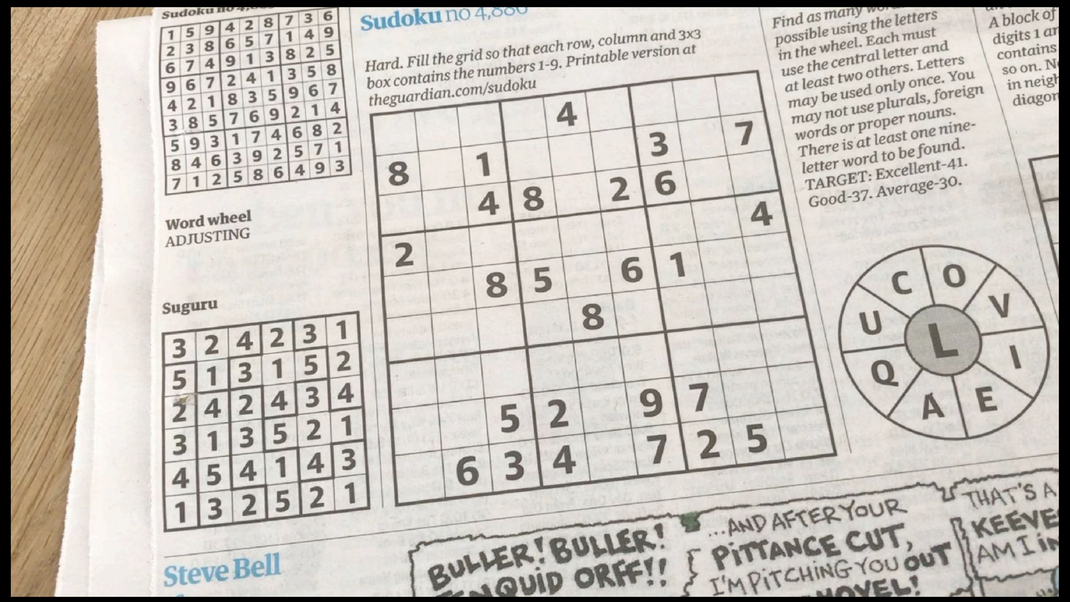
Some Worked Examples

AR Sudoku


Sudoku no 4,880
Hard. Fill the grid so that each row, column and 3x3 box contains the numbers 1-9. Printable version at theguardian.com/sudoku
Word wheel
ADJUSTING
Suguru
Find as many words possible using the letters in the wheel. Each must use the central letter and at least two others. Letters may be used only once. You may not use plurals, foreign words or proper nouns. There is at least one nine-letter word to be found. TARGET: Excellent-41. Good-37. Average-30.
Steve Bell
BULLER! BULLER!
...AND AFTER YOUR PITTANCE CUT, I'M PITCHING YOU OUT

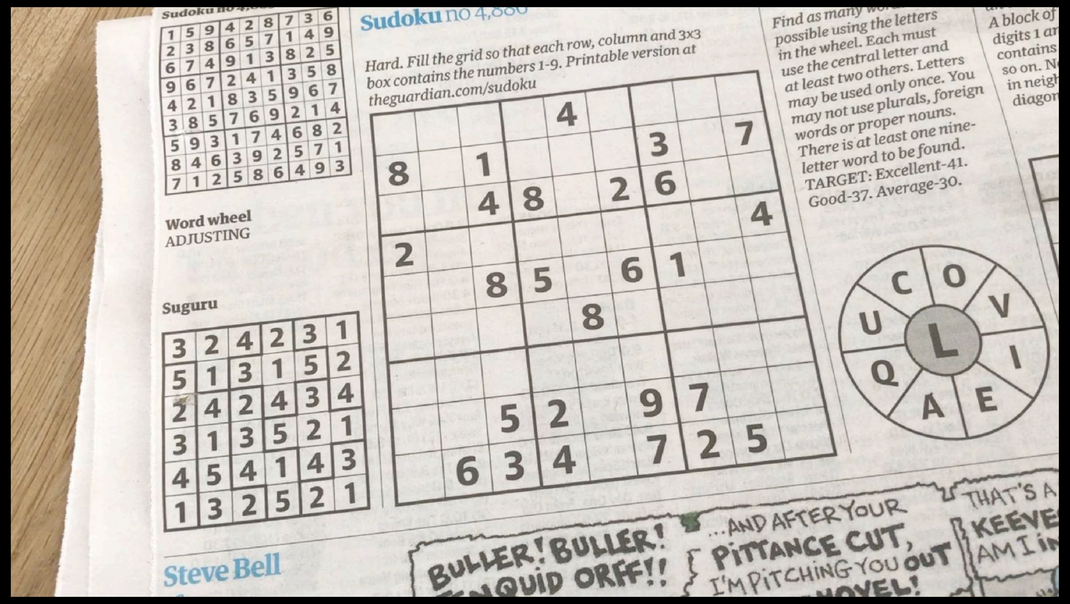
What do we need to do?

- Locate the puzzle
  - Threshold the image
    - Adaptive thresholding
  - Where is the puzzle
    - It’s a square shape?
    - It’s the largest thing in the image!
- Extract the Puzzle
  - We need a homography
    - We need the corners of the puzzle
      - Hough transform?
      - Or cheat?

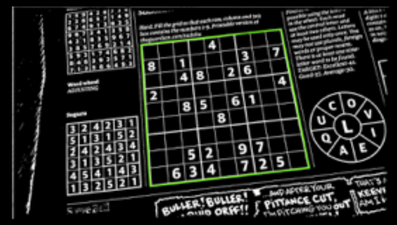


What do we need to do?

- Get the digits
  - Connected component analysis
- OCR
    - Deep learning
- Solve Puzzle
    - Already solved problem - use google



The Pipeline



				4				
8		1				3		7
		4	8		2	6		
2								4
		8	5		6	1		
				8				
		5	2		9	7		
	6	3	4		7	2	5	

				4				
8		1				3		7
		4	8		2	6		
2								4
		8	5		6	1		
				8				
		5	2		9	7		
	6	3	4		7	2	5	

Greyscale/ Threshold → Locate Puzzle → Extract Puzzle → Extract Digits →

→ OCR → Solve → Render

				4				
8		1				3		7
		4	8		2	6		
2								4
		8	5		6	1		
				8				
		5	2		9	7		
	6	3	4		7	2	5	

5	3	6	7	4	1	9	8	2
8	2	1	6	9	5	3	4	7
7	9	4	8	3	2	6	1	5
2	5	9	1	7	3	8	6	4
3	4	8	5	2	6	1	7	9
6	1	7	9	8	4	5	2	3
1	7	2	3	5	8	4	9	6
4	8	5	2	6	9	7	3	1
9	6	3	4	1	7	2	5	8

Sudoku Grab - Convert to Greyscale

Sudoku no 4,880

Hard. Fill the grid so that each row, column and 3x3 box contains the numbers 1-9. Printable version at theguardian.com/sudoku

				4				
8		1				3		7
		4	8		2	6		
2								4
		8	5		6	1		
				8				
		5	2		9	7		
	6	3	4		7	2	5	

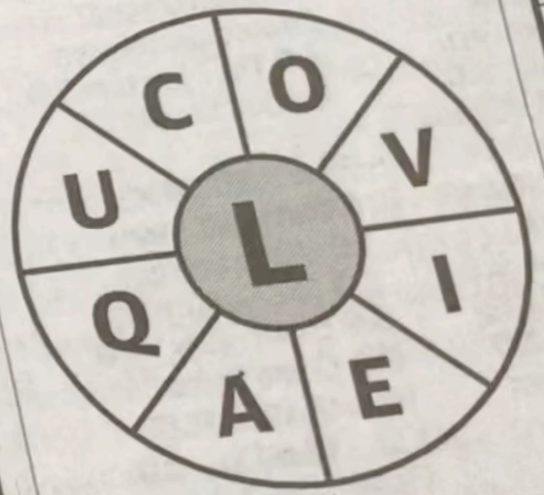
Sudoku no 4,880

2	3	8	6	5	4	1	7	9
6	7	4	9	1	3	8	2	5
9	6	7	2	4	1	3	5	8
4	2	1	8	3	5	9	6	7
3	8	5	7	6	9	2	1	4
5	9	3	1	7	4	6	8	2
8	4	6	3	9	2	5	7	1
7	1	2	5	8	6	4	9	3

Word wheel

ADJUSTING

Find as many words possible using the letters in the wheel. Each must use the central letter and at least two others. Letters may be used only once. You may not use plurals, foreign words or proper nouns. There is at least one nine-letter word to be found. TARGET: Excellent-41. Good-37. Average-30.

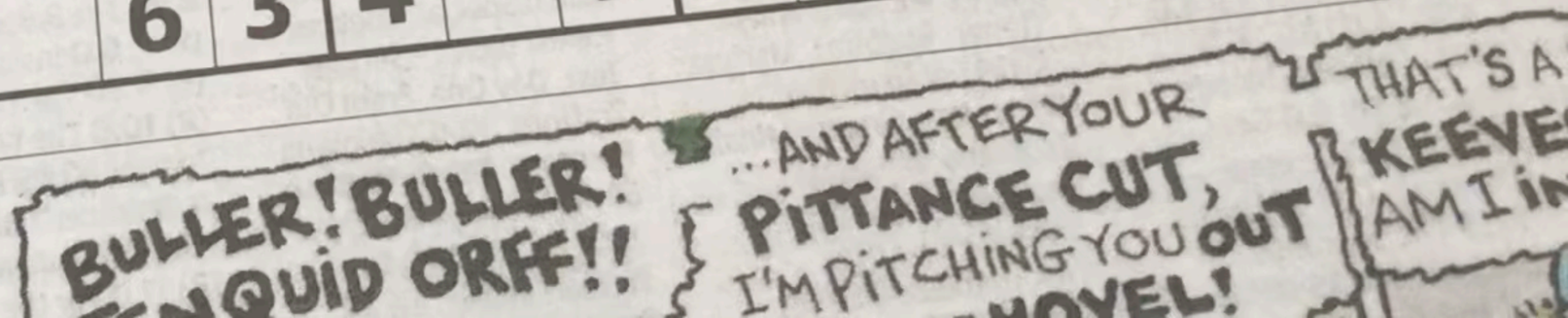


Suguru

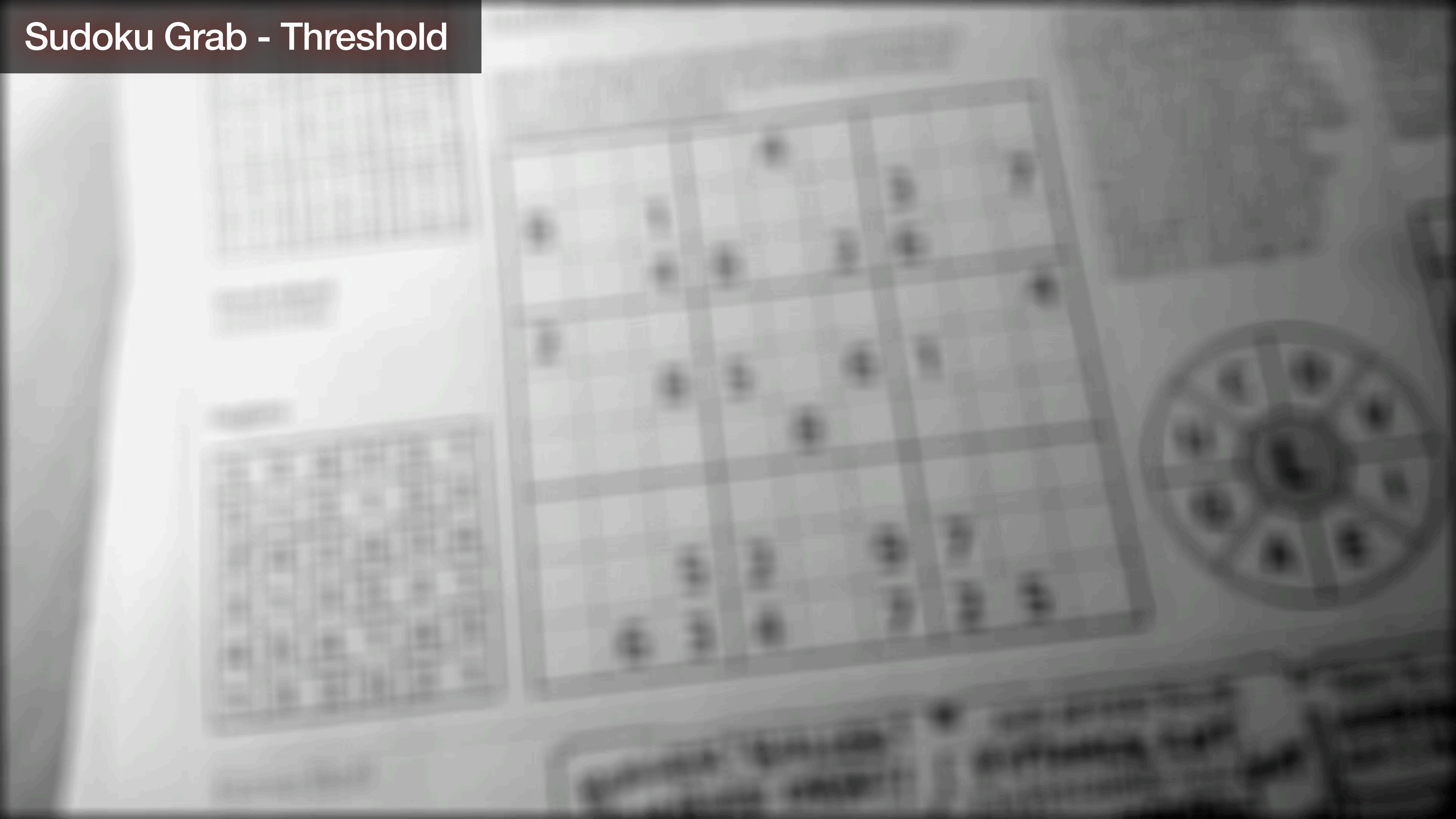
3	2	4	2	3	1
5	1	3	1	5	2
2	4	2	4	3	4
3	1	3	5	2	1
4	5	4	1	4	3
1	3	2	5	2	1

A block of digits 1 an contains so on. N in neigh diagon

Steve Bell



Sudoku Grab - Threshold



Sudoku Grab - Locate Puzzle

6	7	4	9	1	3	8	2	5
9	6	7	2	4	1	3	5	8
4	2	1	8	3	5	9	6	7
3	8	5	7	6	9	2	1	4
5	9	3	1	7	4	6	8	2
8	4	6	3	9	2	5	7	1
7	1	2	5	8	6	4	9	3

Word wheel
ADJUSTING

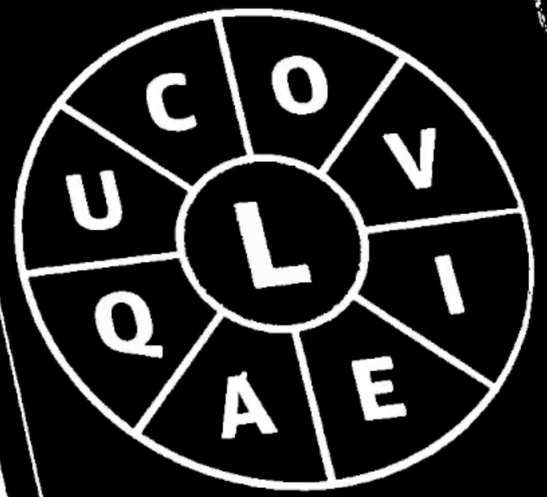
Suguru

3	2	4	2	3	1
5	1	3	1	5	2
2	4	2	4	3	4
3	1	3	5	2	1
4	5	4	1	4	3
1	3	2	5	2	1

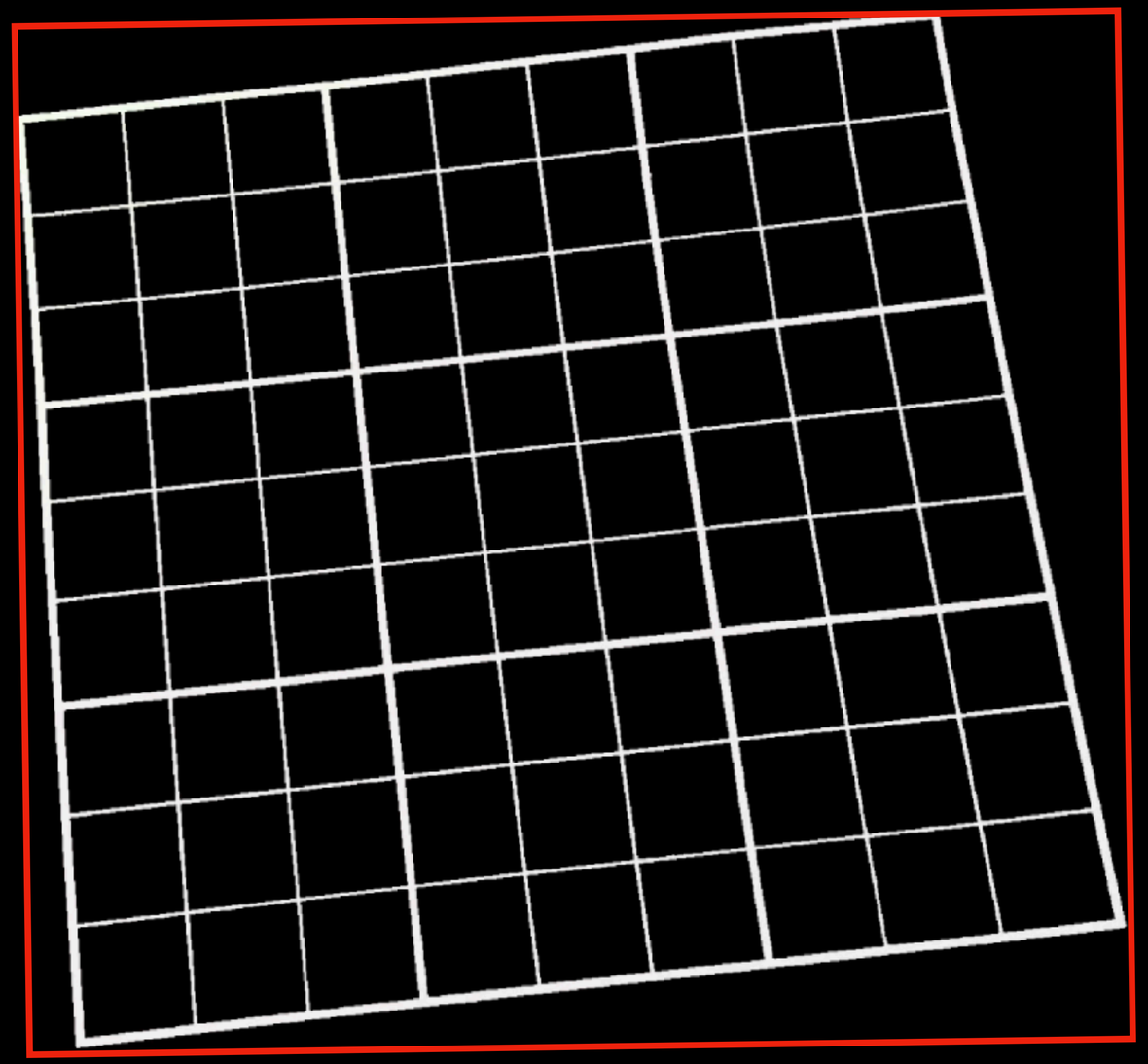
Hard. Fill the grid so that each row, column and 3x3 box contains the numbers 1-9. Printable version at theguardian.com/sudoku

				4				
						3		7
8		1						
		4	8		2	6		
								4
2								
		8	5		6	1		
				8				
		5	2		9	7		
	6	3	4		7	2	5	

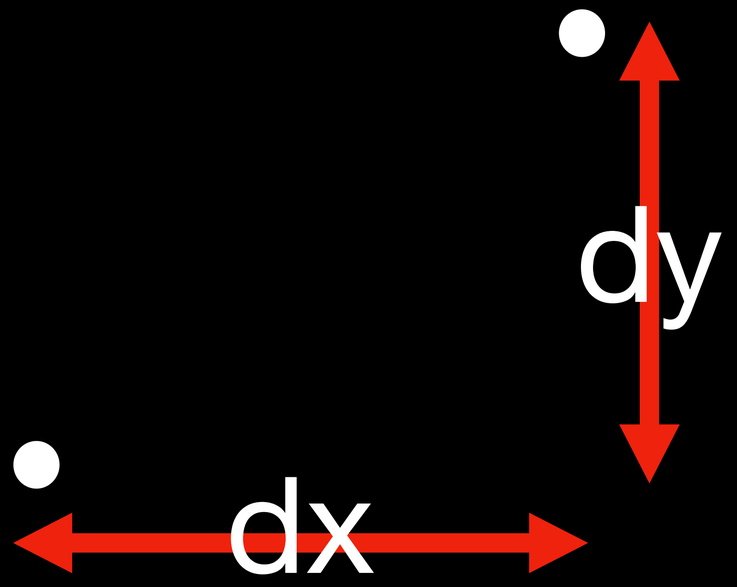
possible using the letters in the wheel. Each must use the central letter and at least two others. Letters may be used only once. You may not use plurals, foreign words or proper nouns. There is at least one nine-letter word to be found. TARGET: Excellent-41. Good-37. Average-30.



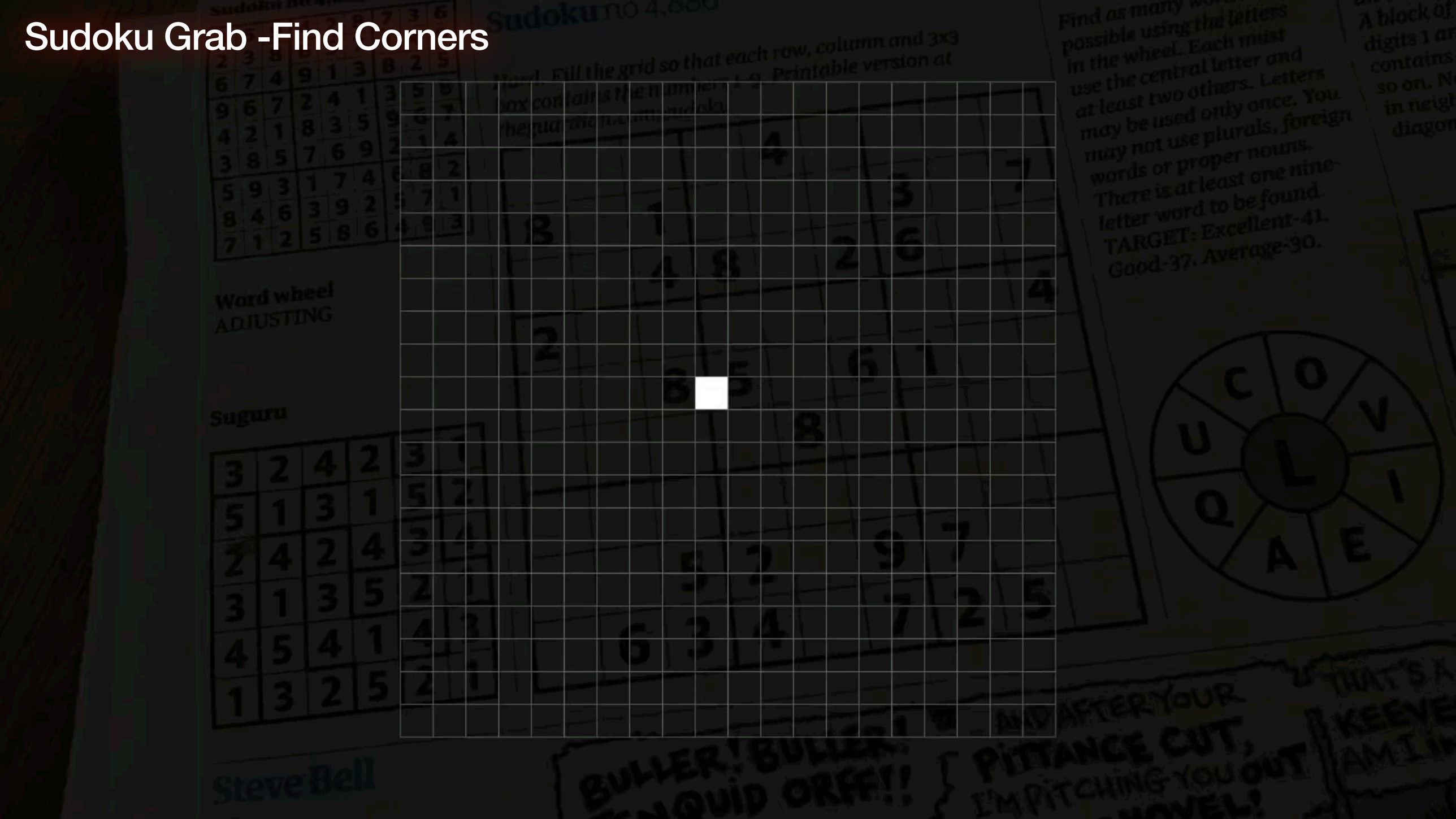
Sudoku Grab -Find Corners



$$d = dx + dy$$



Sudoku Grab -Find Corners



Sudoku Grab -Extract Puzzle

Sudoku no 4,880

2	3	8	6	5	4	7	1	9
6	7	4	9	1	3	8	2	5
9	6	7	2	4	1	3	5	8
4	2	1	8	3	5	9	6	7
3	8	5	7	6	9	2	1	4
5	9	3	1	7	4	6	8	2
8	4	6	3	9	2	5	7	1
7	1	2	5	8	6	4	9	3

Word wheel
ADJUSTING

Suguru

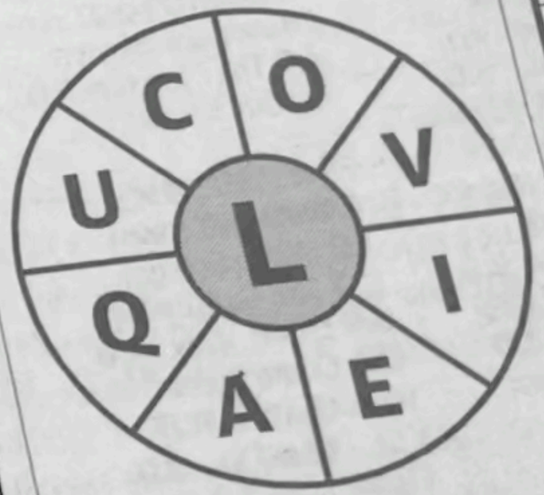
3	2	4	2	3	1
5	1	3	1	5	2
2	4	2	4	3	4
3	1	3	5	2	1
4	5	4	1	4	3
1	3	2	5	2	1

Sudoku no 4,880

Hard. Fill the grid so that each row, column and 3x3 box contains the numbers 1-9. Printable version at theguardian.com/sudoku

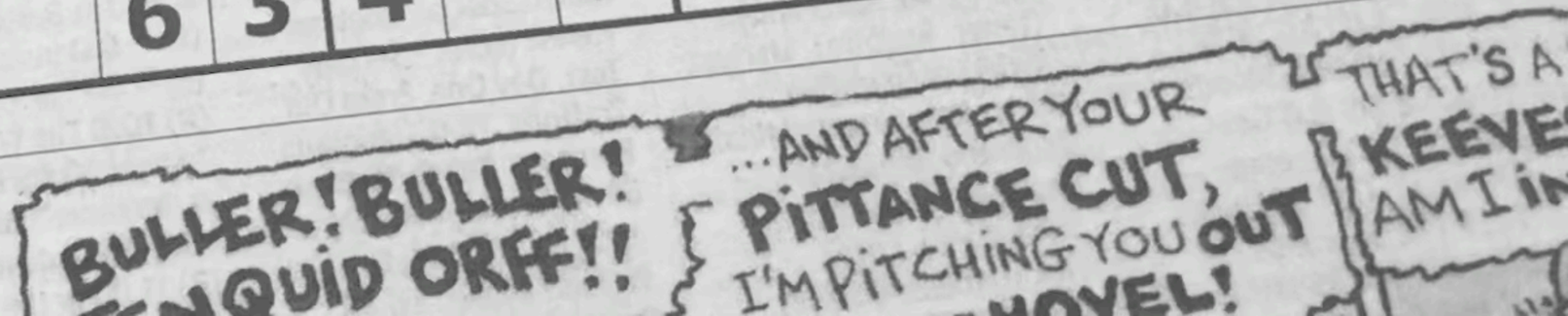
				4				
						3		7
8		1						
		4	8		2	6		
								4
2								
		8	5		6	1		
				8				
		5	2		9	7		
	6	3	4		7	2	5	

Find as many words possible using the letters in the wheel. Each must use the central letter and at least two others. Letters may be used only once. You may not use plurals, foreign words or proper nouns. There is at least one nine-letter word to be found. TARGET: Excellent-41. Good-37. Average-30.

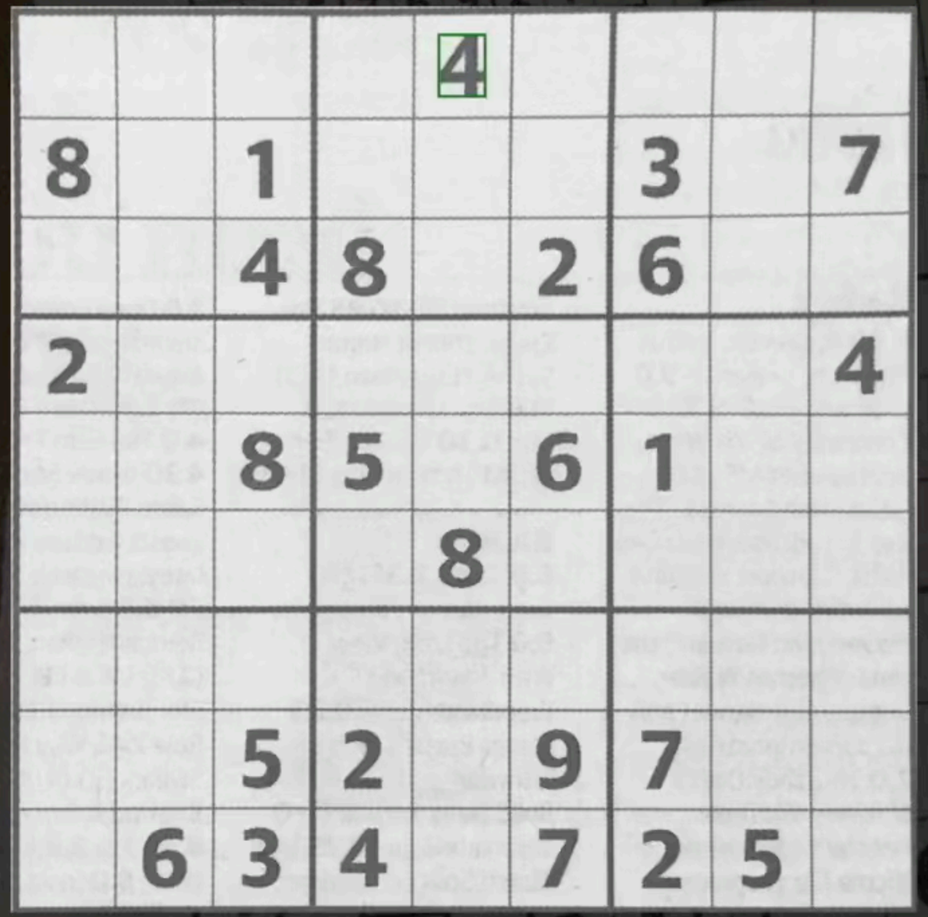


A block of
digits 1 an
contains
so on. N
in neigh
diagon

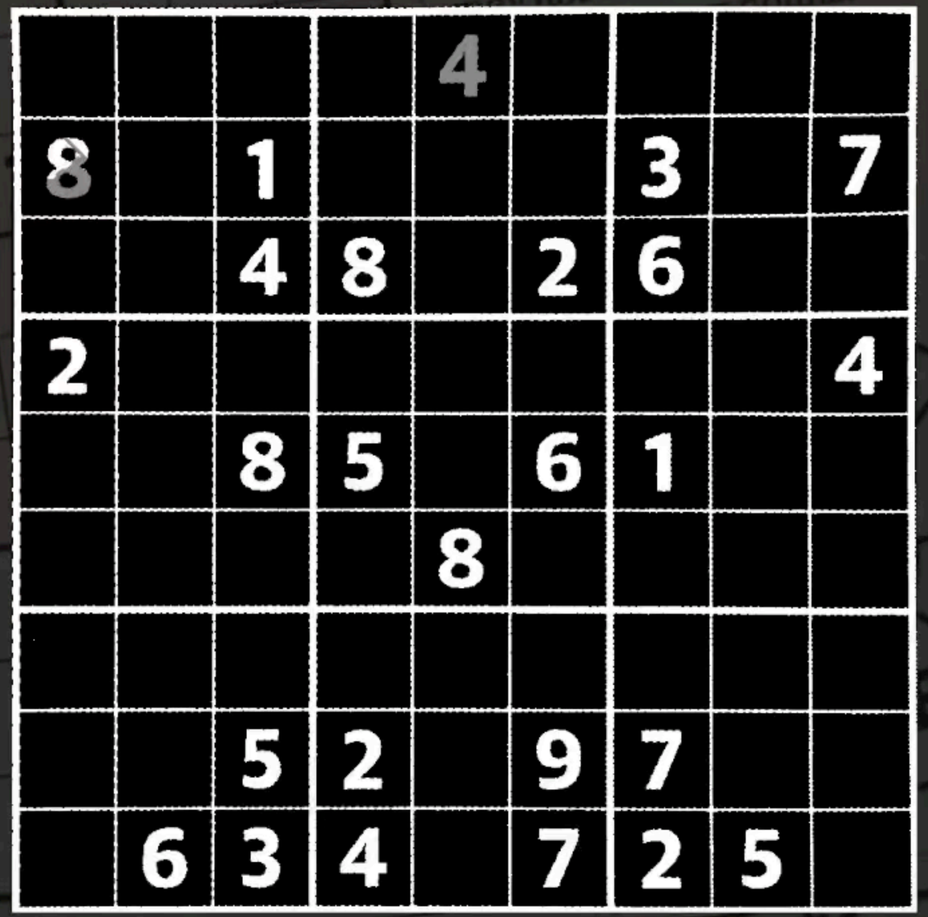
Steve Bell



Sudoku Grab -Extract Digits

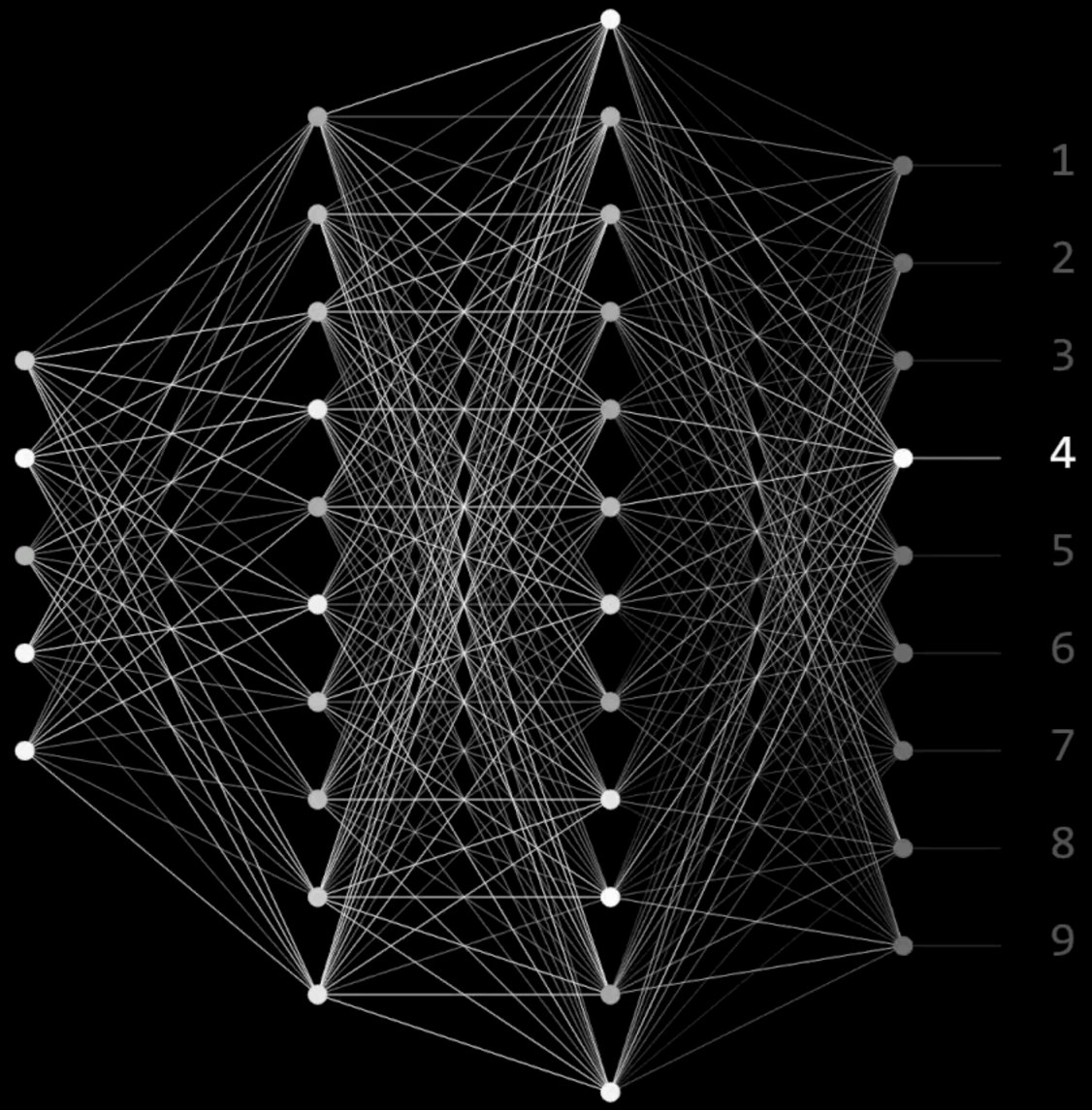


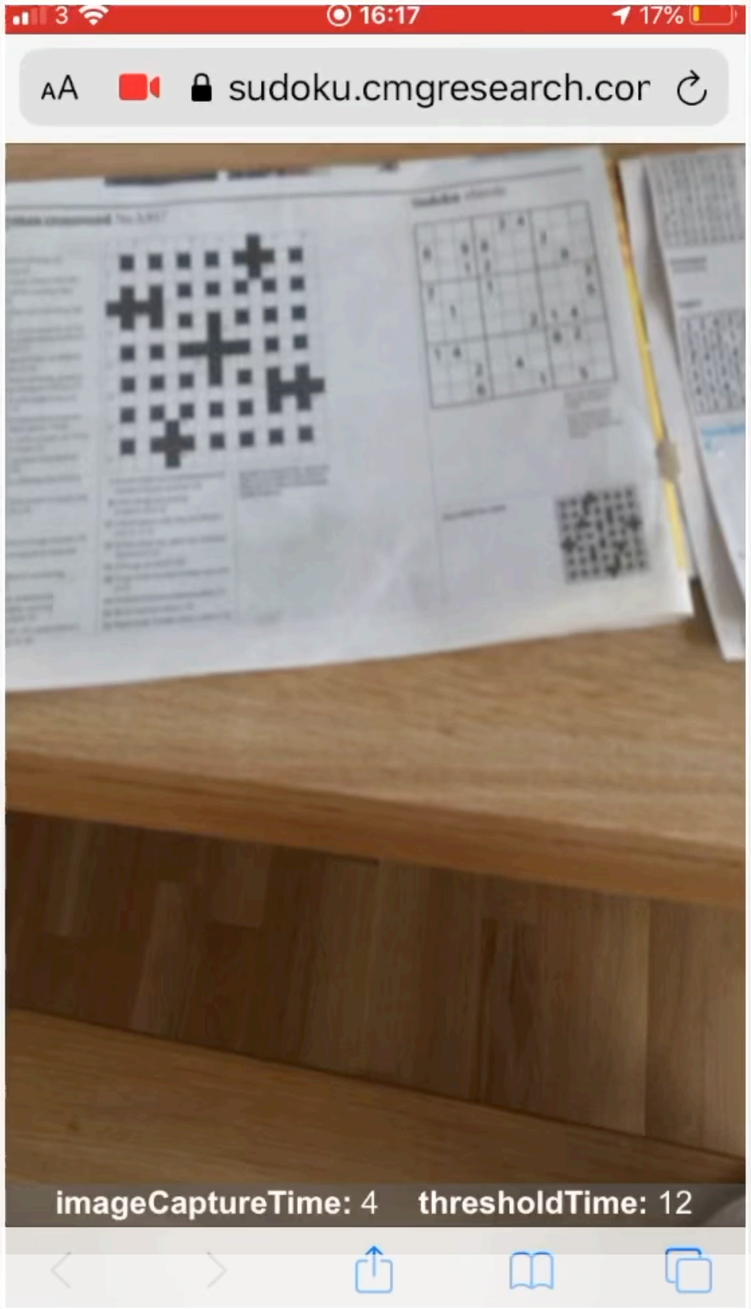
				4				
8		1				3		7
		4	8		2	6		
2								4
		8	5		6	1		
				8				
		5	2		9	7		
	6	3	4		7	2	5	



				4				
8		1				3		7
		4	8		2	6		
2								4
		8	5		6	1		
				8				
		5	2		9	7		
	6	3	4		7	2	5	

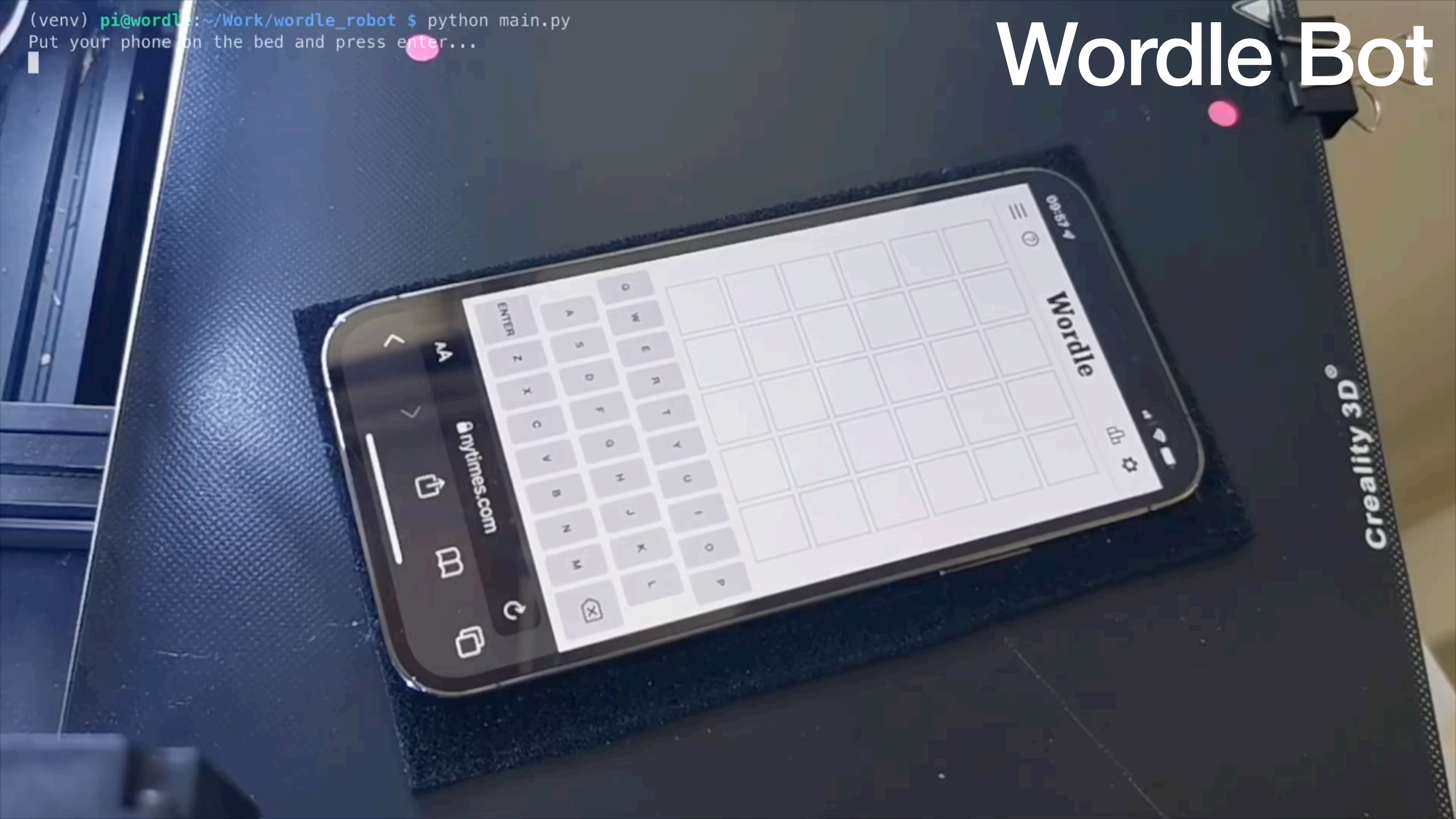
Sudoku Grab -OCR





https://sudoku.cmgresearch.com/

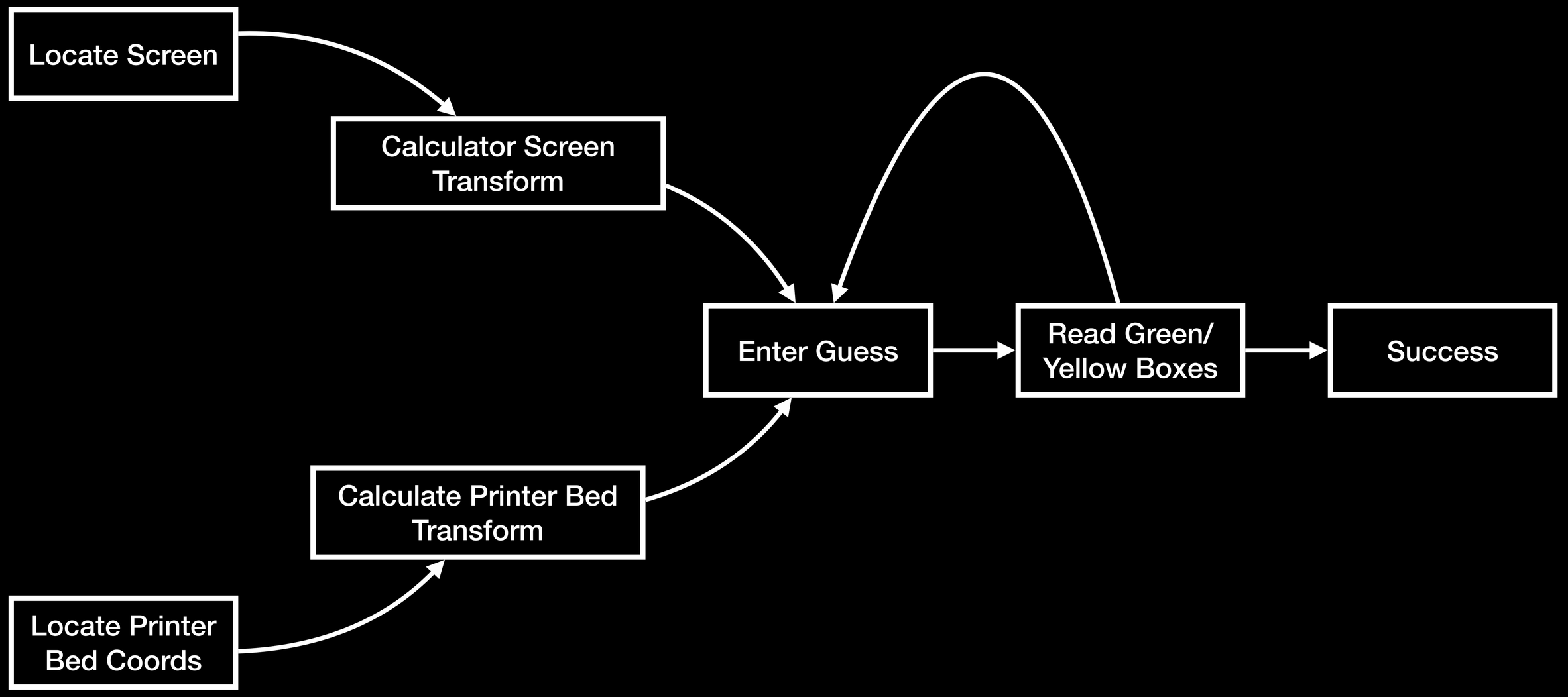
Wordle Bot


(venv) pi@wordle:~/Work/wordle_robot $ python main.py
Put your phone on the bed and press enter...
Wordle
nytimes.com
Creality 3D

What are the problems to solve?

- Calibrate robot
- Locate the phone’s screen
- Read the result box colours
- Predict the next best guess

Pipeline


Locate Screen
Calculator Screen Transform
Enter Guess
Read Green/ Yellow Boxes
Success
Calculate Printer Bed Transform
Locate Printer Bed Coords

Locate Screen

Otsu Threshold → Connected Component Analysis → Matched Rectangle → Affine Transform

Printer Bed Extents

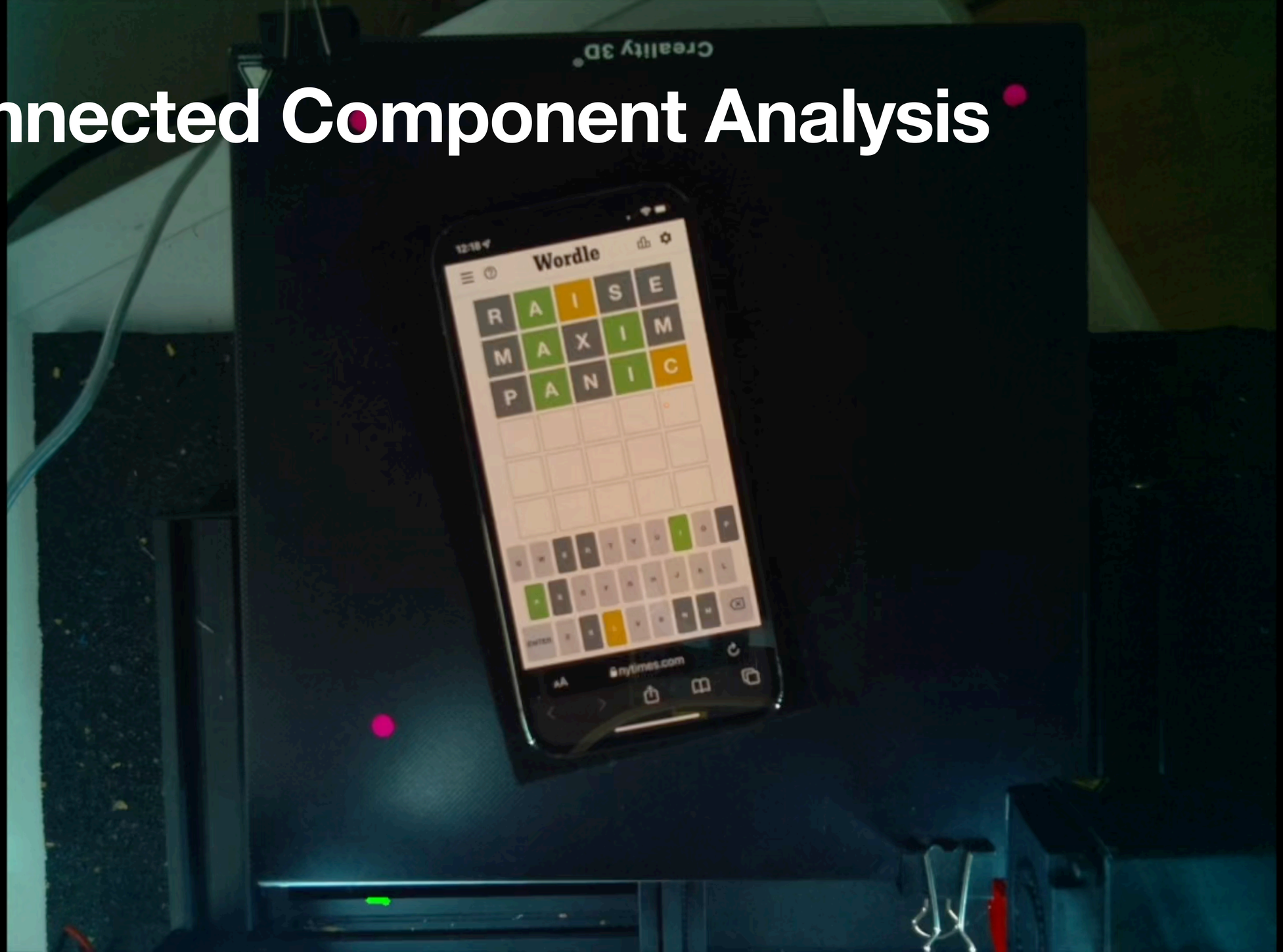
HSV Segmentation → Connected Component Analysis → Matched Magenta Circles → Affine Transform

Letter Colours

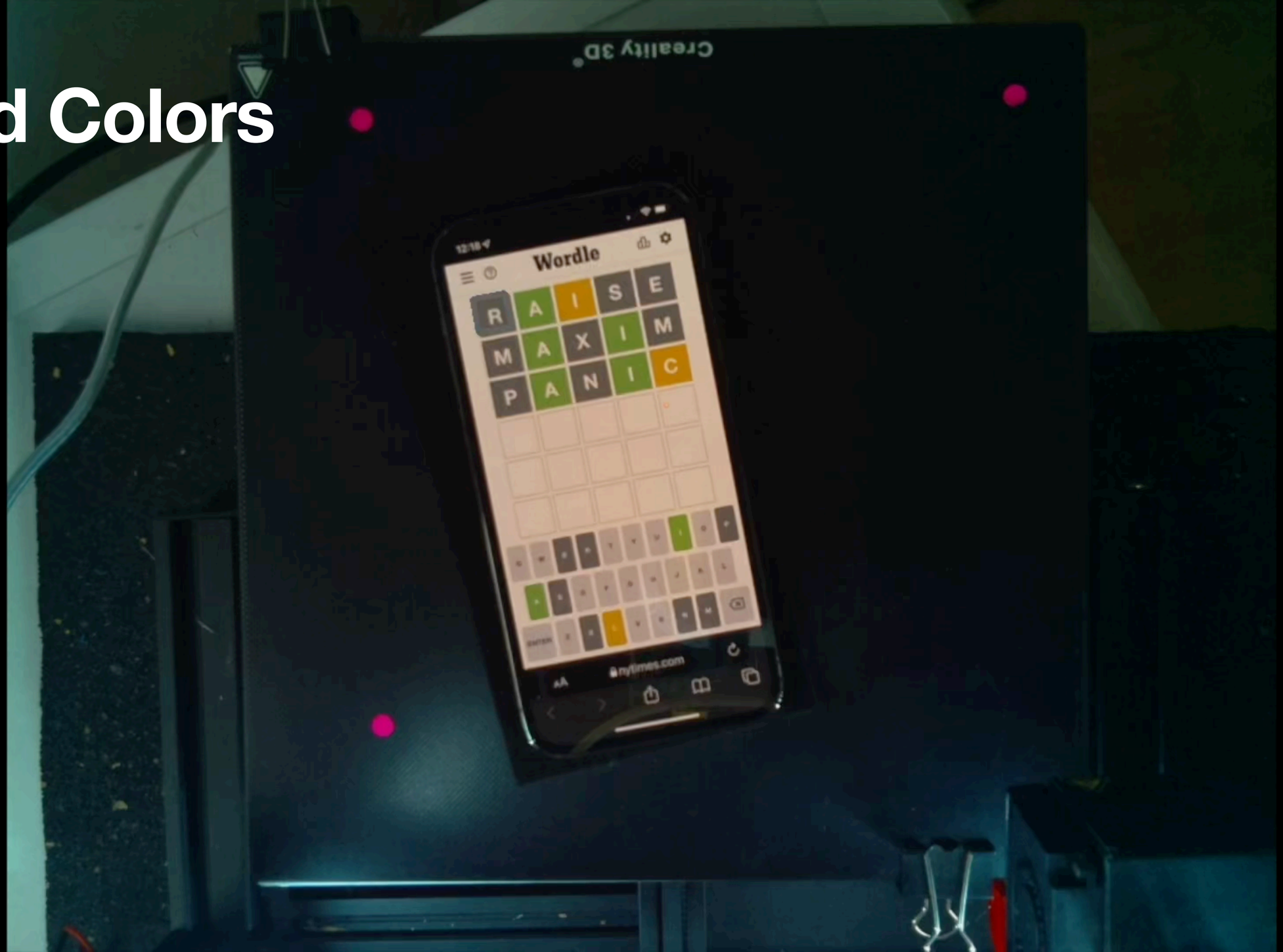
HSV Segmentation

Otsu Threshold

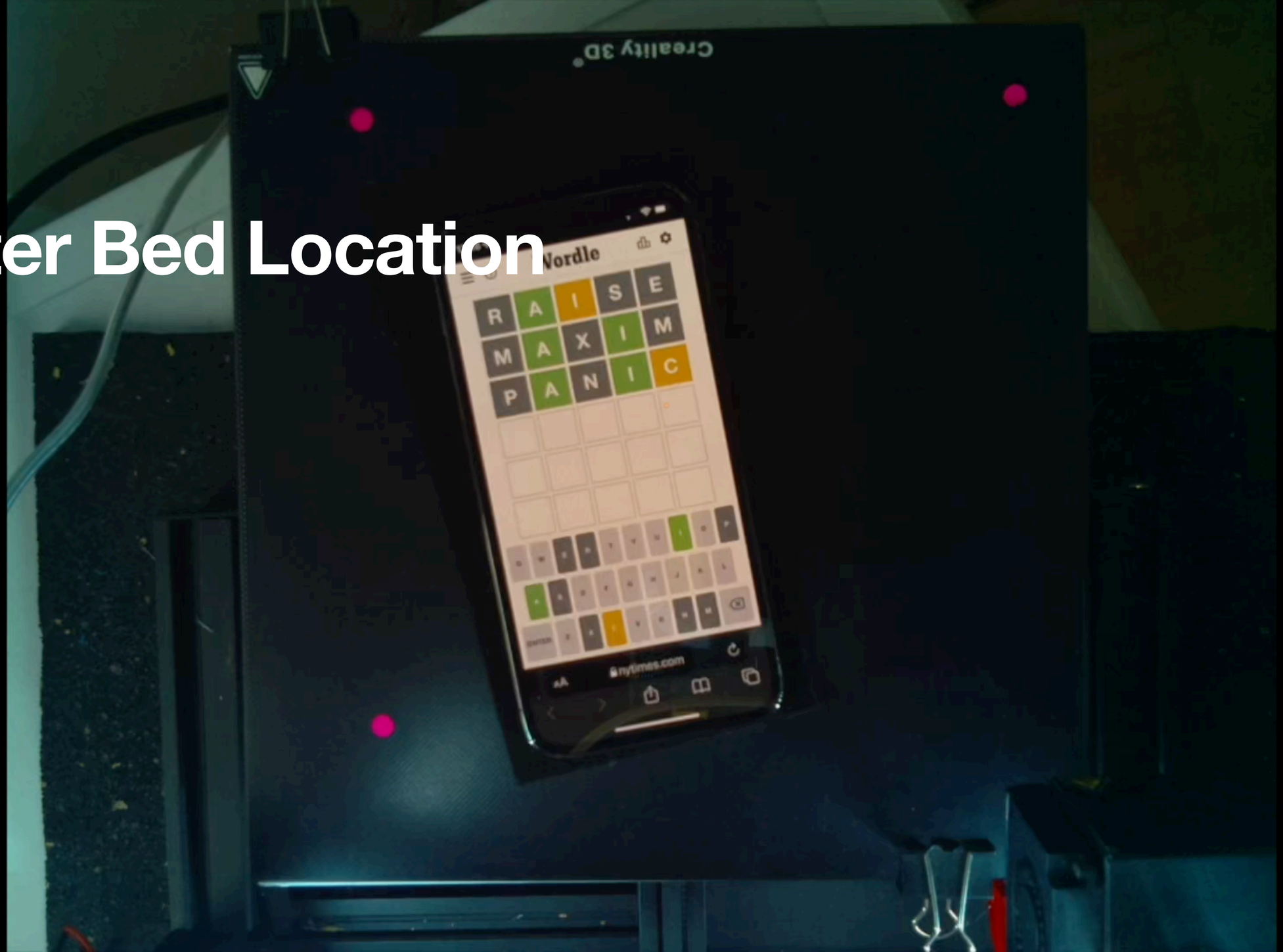
Connected Component Analysis

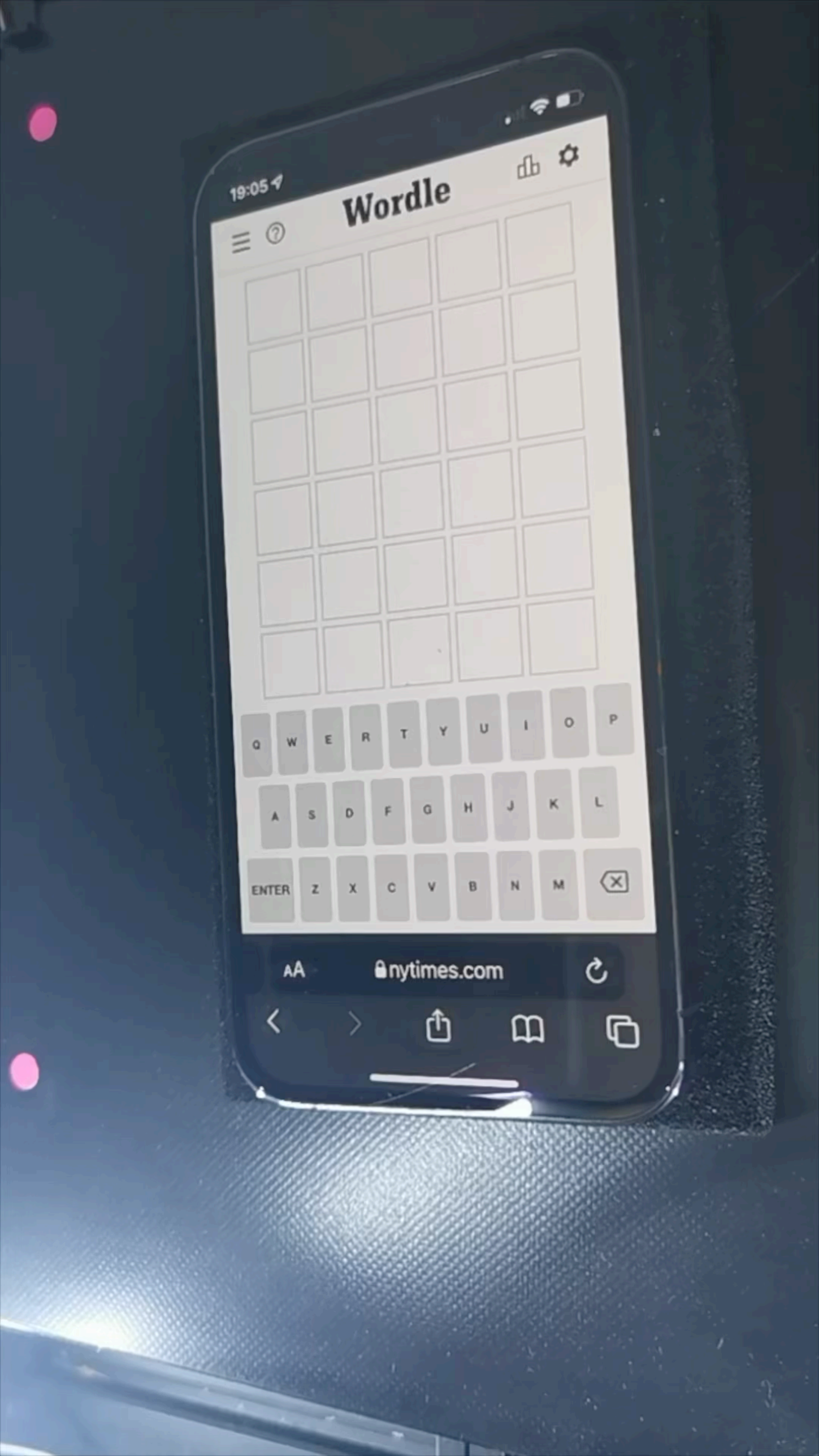


Grid Colors



Printer Bed Location




19:05
Wordle
Q W E R T Y U I O P
A S D F G H J K L
ENTER Z X C V B N M
AA
nytimes.com