COGS300

Deep Learning

Instructor: Márton Sóskuthy
marton.soskuthy@ubc.ca

TAs: Daichi Furukawa · Victoria Lim · Amy Wang
cogs.300@ubc.ca

UBC
Human
AI
Interaction

Participants needed for a user-study

Explainable AI in Intelligent Tutoring Systems.

Looking for participants who are majoring in Computer Science or related fields and meet the following requirements:

- HAVE taken CPSC 110.
- HAVE NOT taken or ARE CURRENTLY taking CPSC 322.
- do not have color vision deficiency.

Max upto 3 Hours

Typically 2 to 2.5 hours

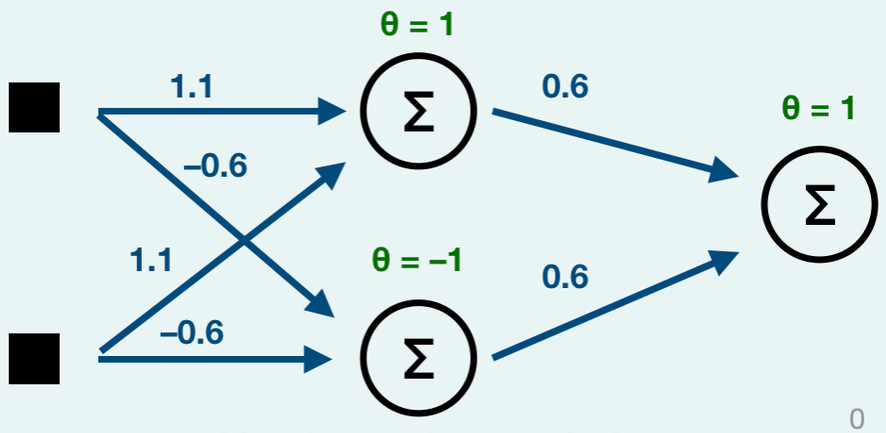
ICICS/CS Building

Get $40

Sign Up:

https://hai.cs.ubc.ca/xai_study

implements



implements

Connectionist system implemented by symbol system, and implementing symbol system

XOR

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

$$p \oplus q = (p \vee q) \wedge \neg(p \wedge q)$$

Is this a symbol system or a connectionist system?

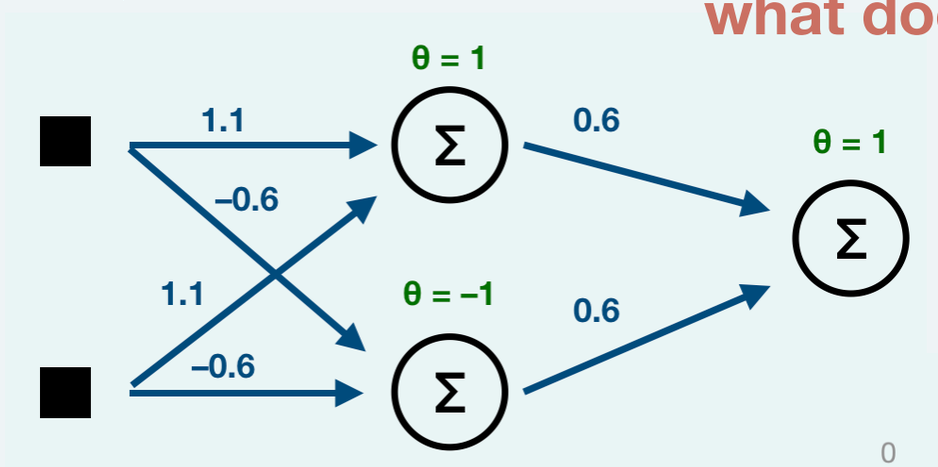
Marr’s levels of analysis

implementational level

physical implementation?

algorithmic/representational level

representations / processes?



computational level

what does the system do? goals?

XOR

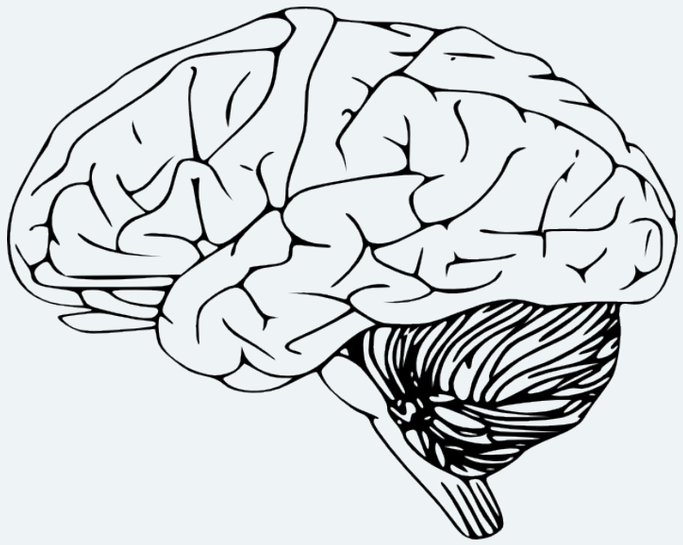
0	0	0
0	1	1
1	0	1
1	1	0

$$p \oplus q = (p \vee q) \wedge \neg(p \wedge q)$$

Marr's levels of analysis

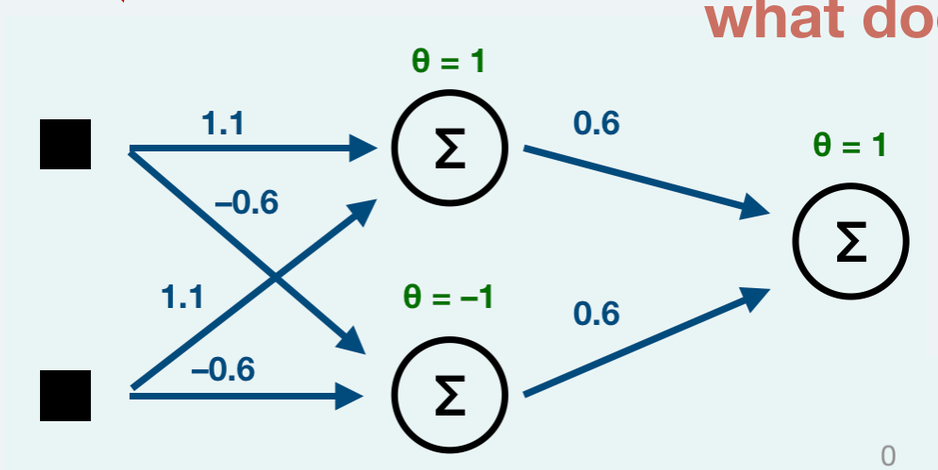
implementational level

physical implementation?



algorithmic/representational level

representations / processes?



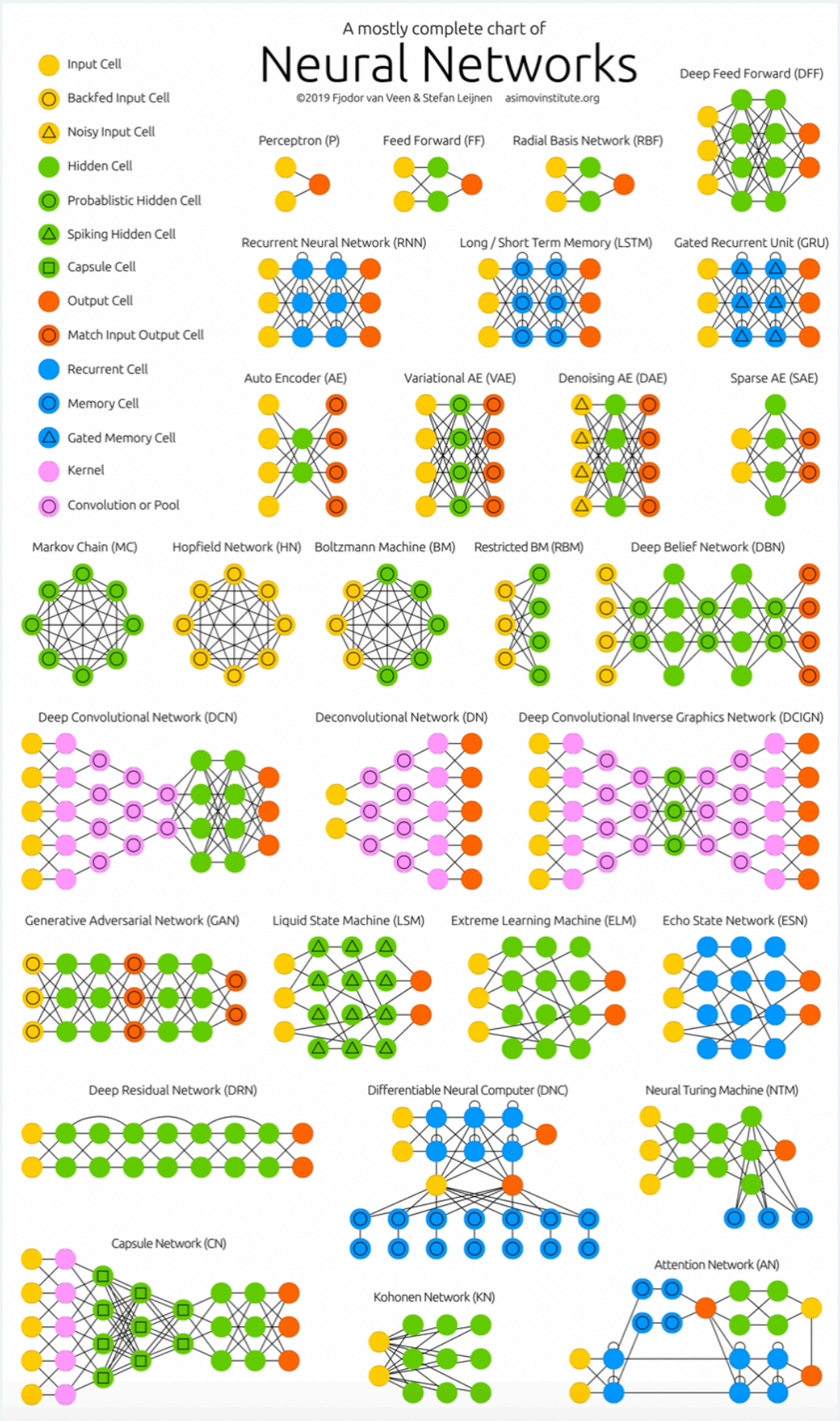
computational level

what does the system do? goals?

XOR

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

$$p \oplus q = (p \vee q) \wedge \neg(p \wedge q)$$

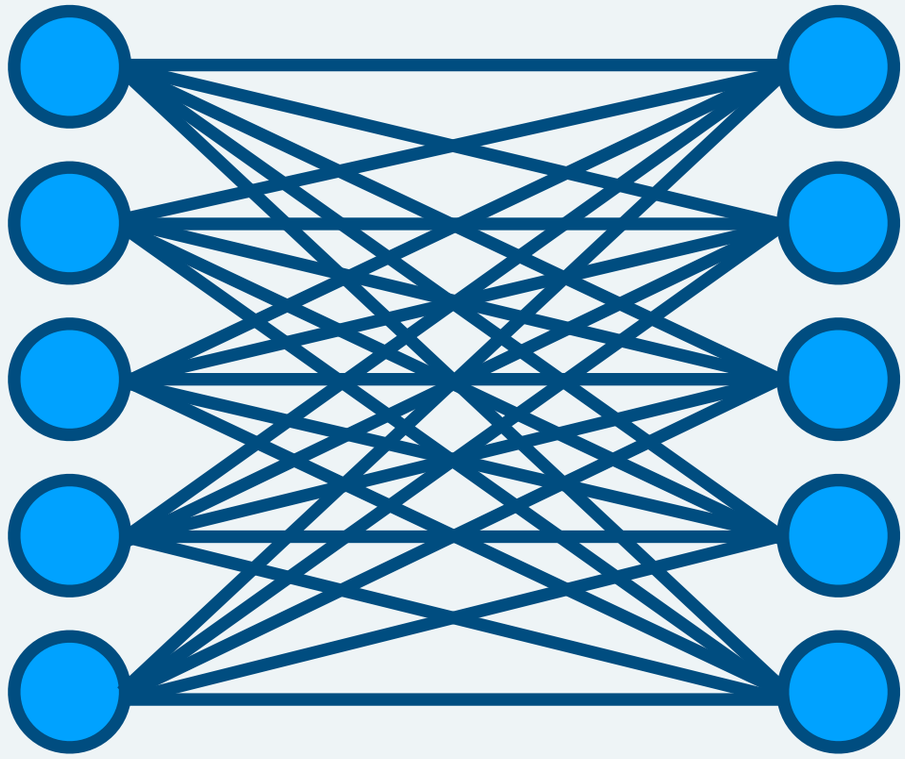


Connectionist networks (2019)

https://www.asimovinstitute.org/neural-network-zoo/

Connectionist networks

Fully Connected (Feed Forward)



all going in same direction, not skipping layers

Connectionist networks

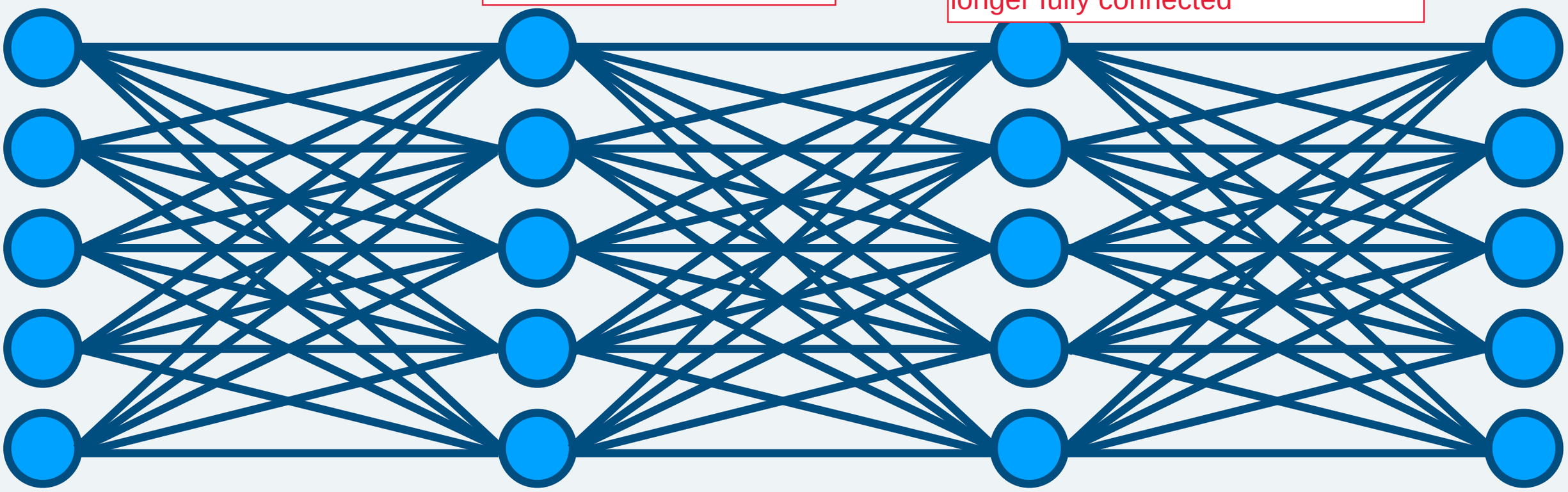
Sequential

Deep Fully Connected (Feed Forward)

Loops = recurrent NN

Typically deep networks have at least 3 layers

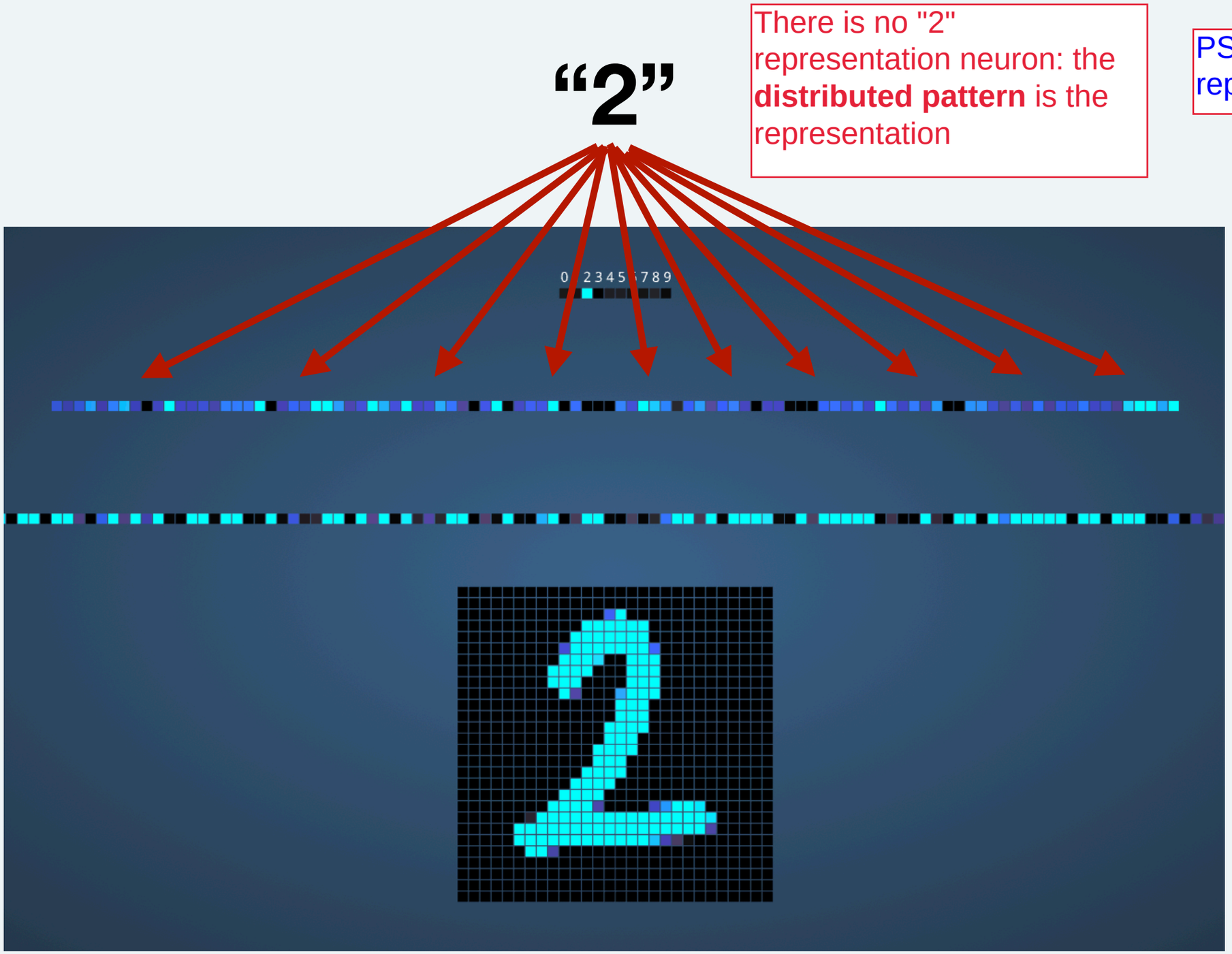
Often NNs aren't fully connected (just dropping 1 connection --> no longer fully connected



Connectionist networks

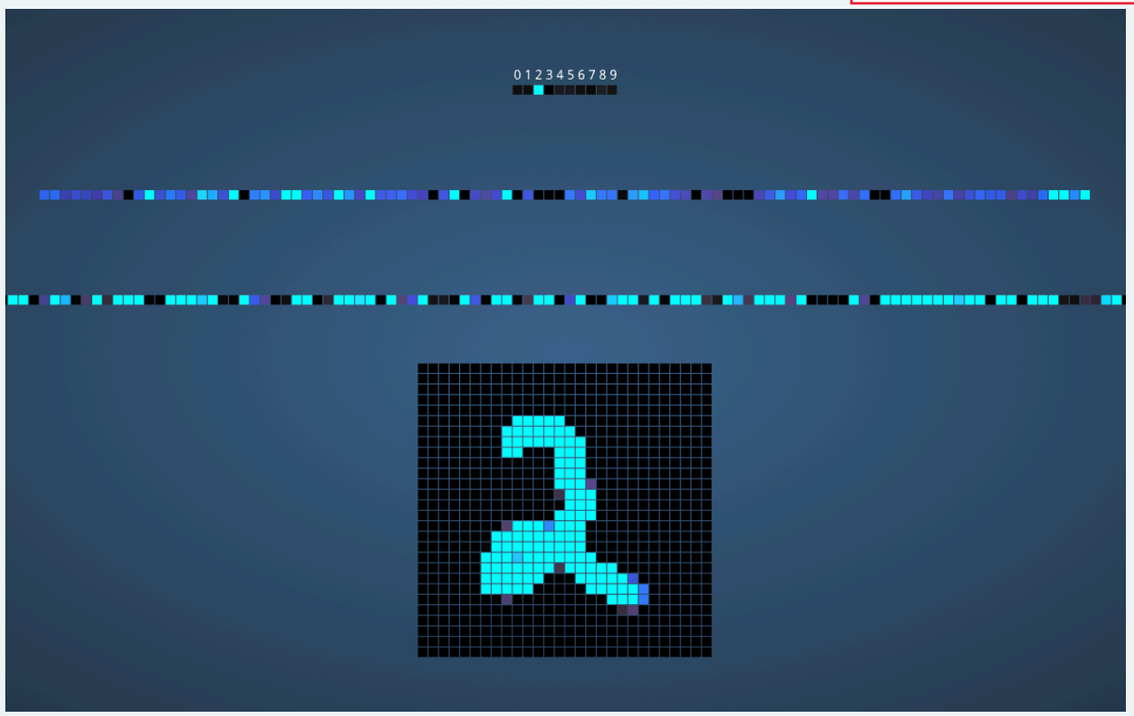
https://adamharley.com/nn_vis/

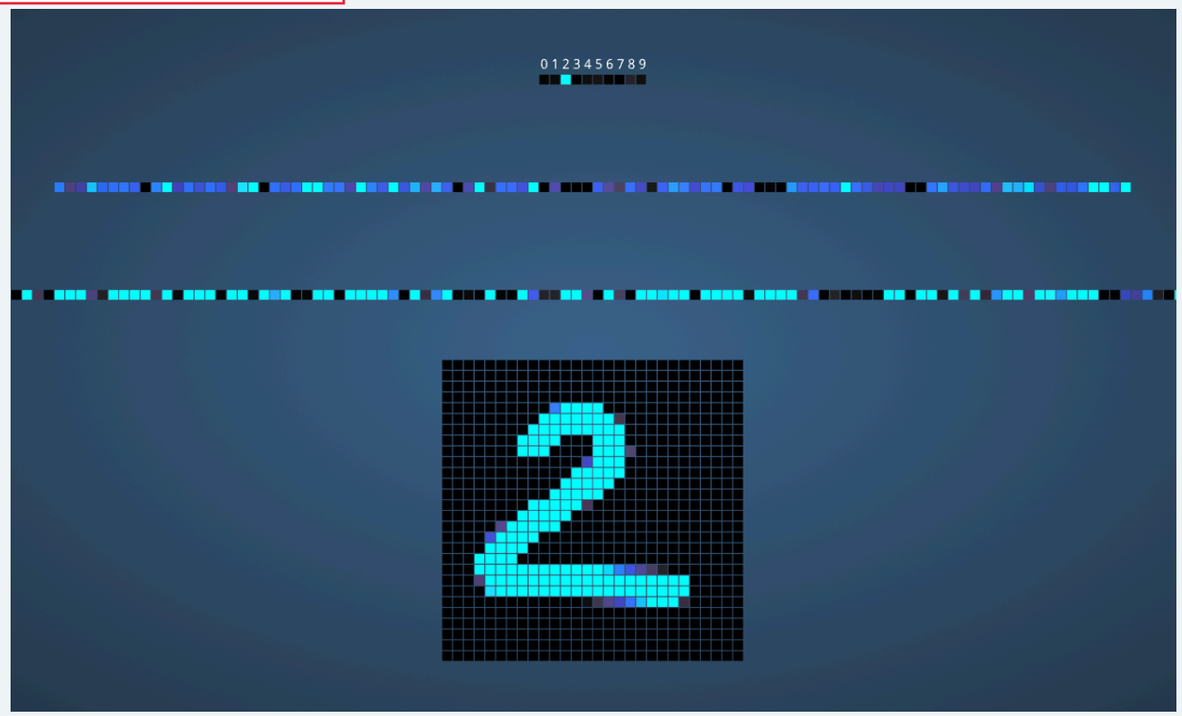
Distributed representations

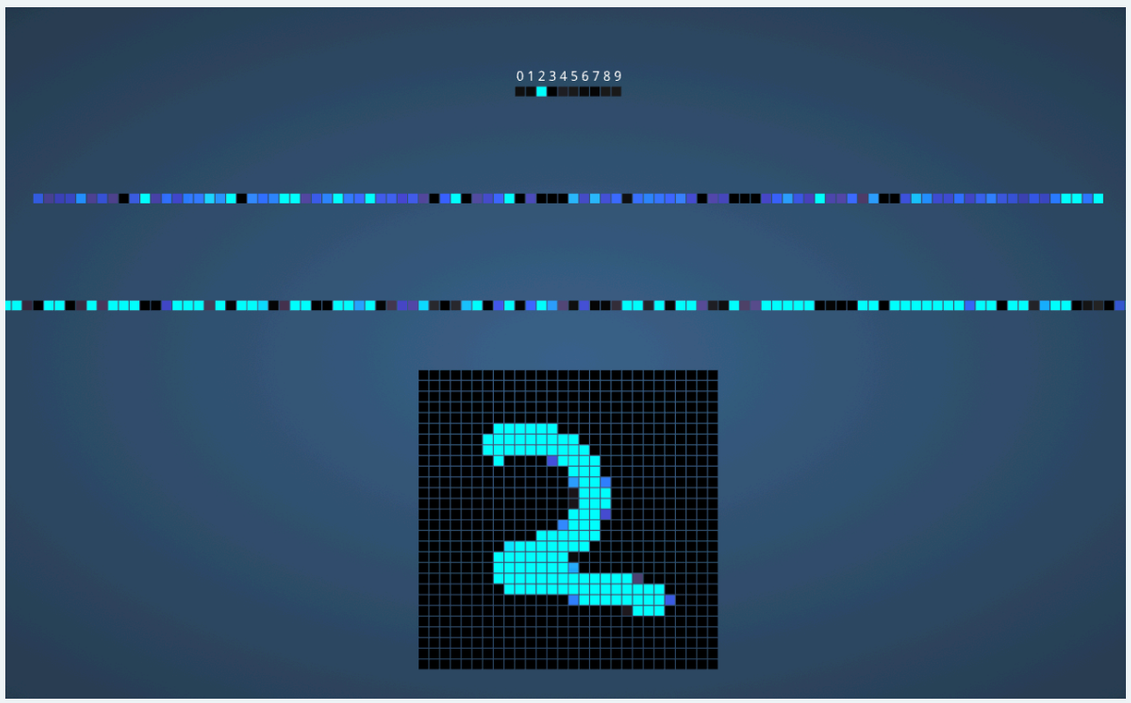


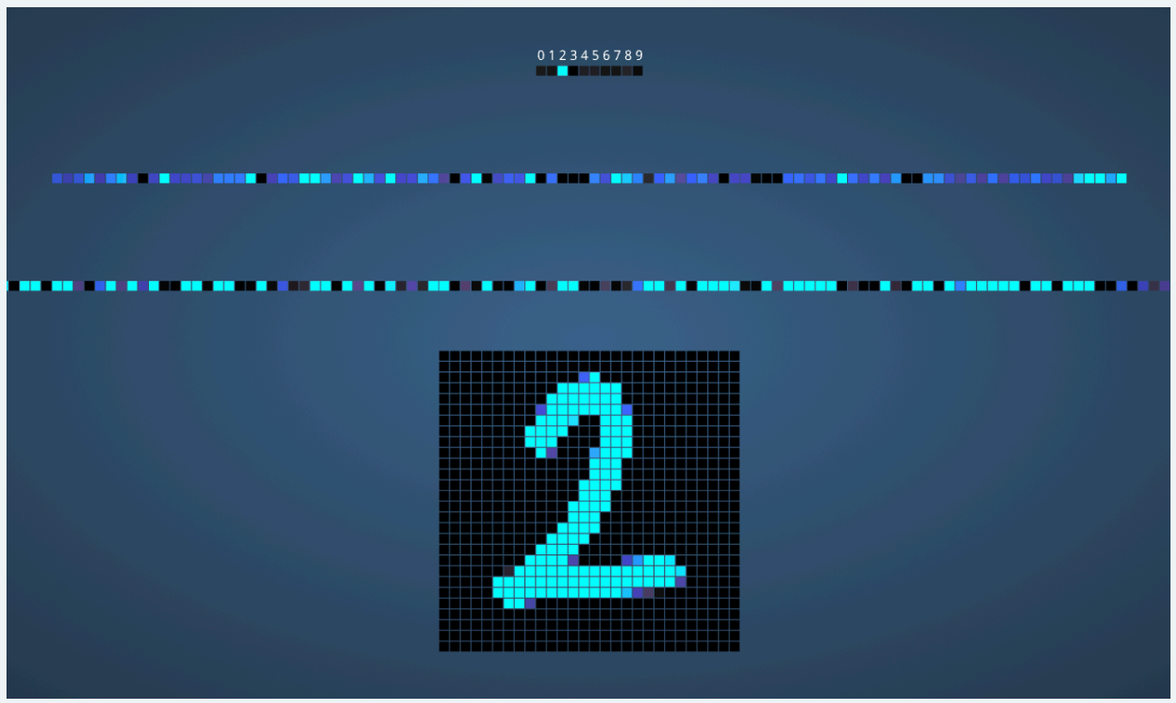
Superpositional representations

all similar activation patterns

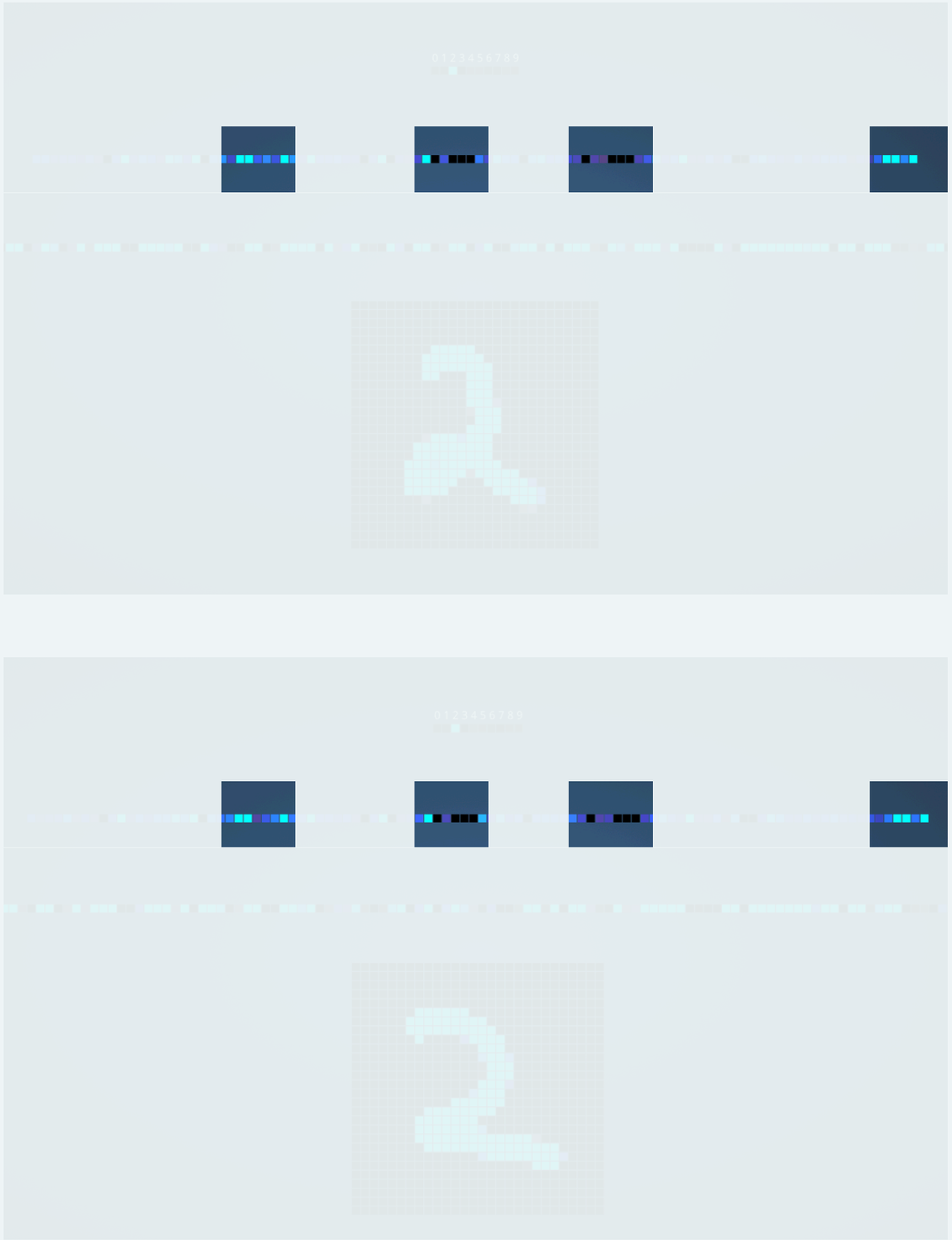


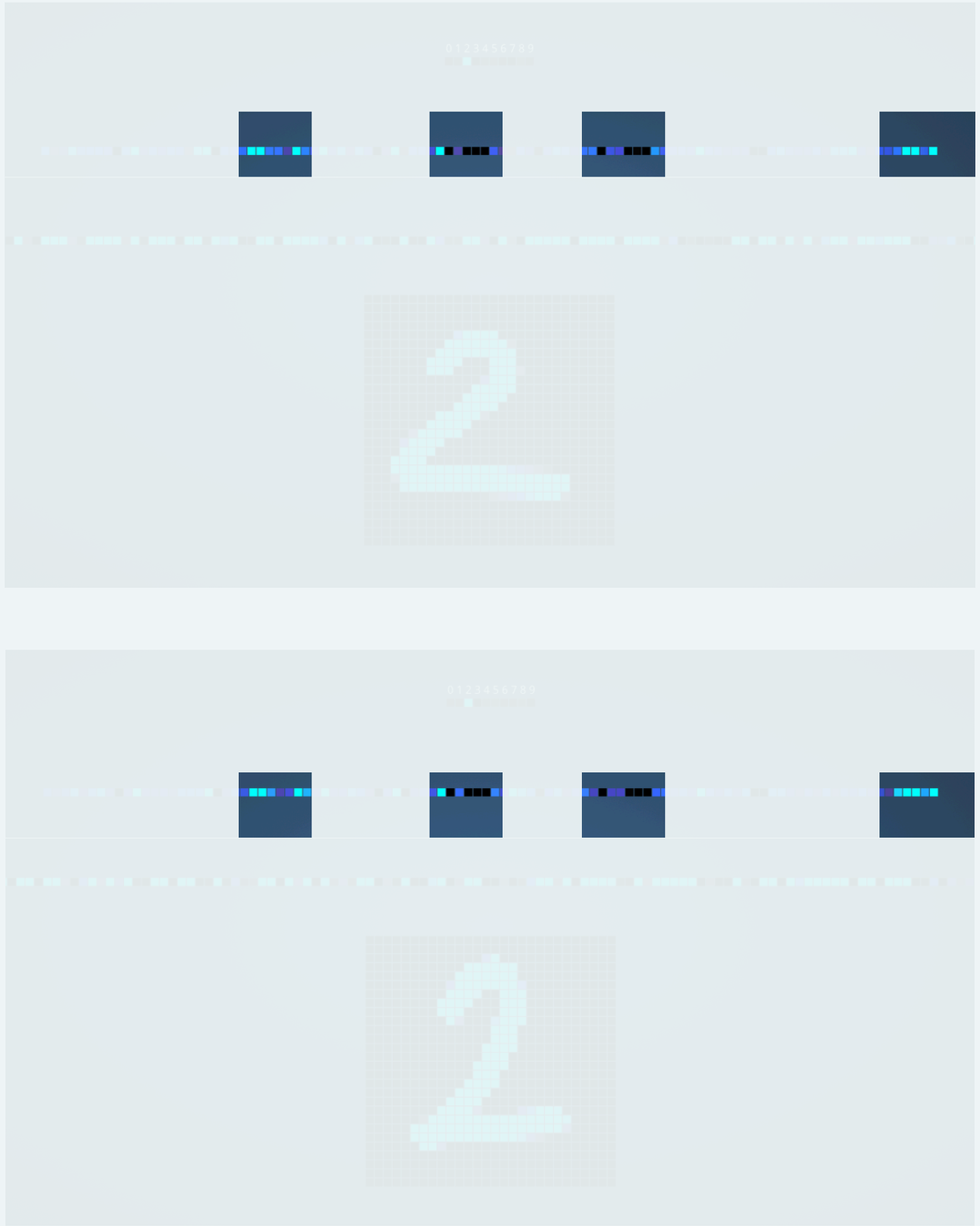






Superpositional representations



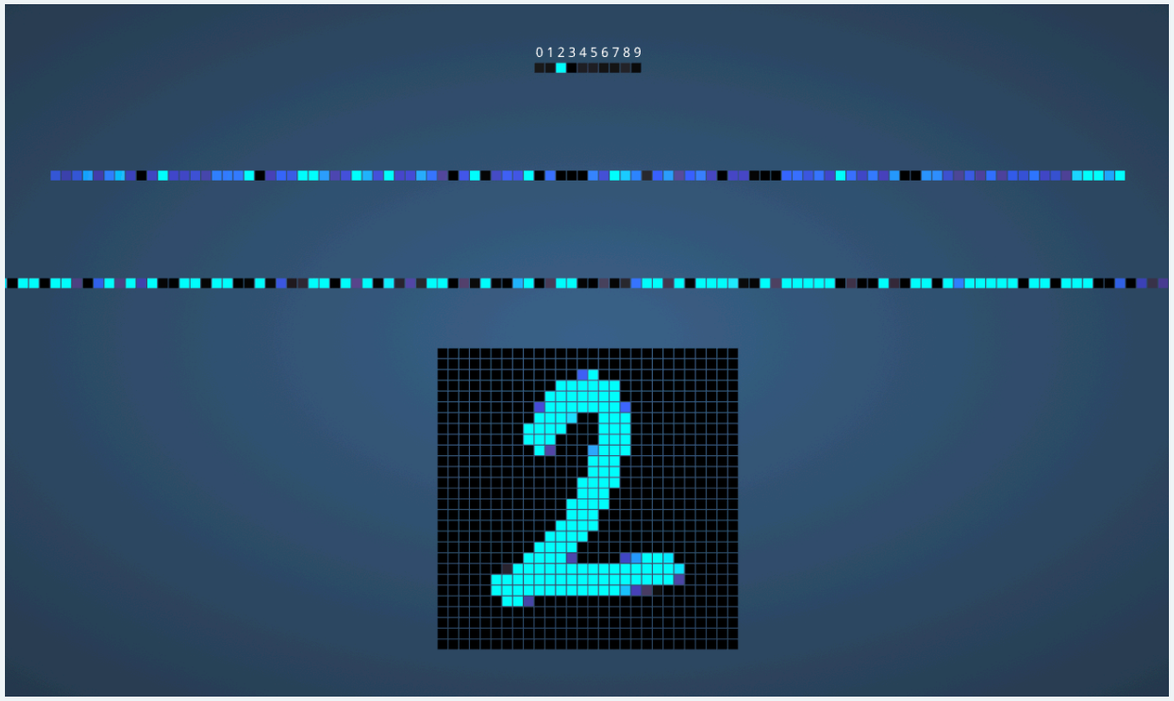


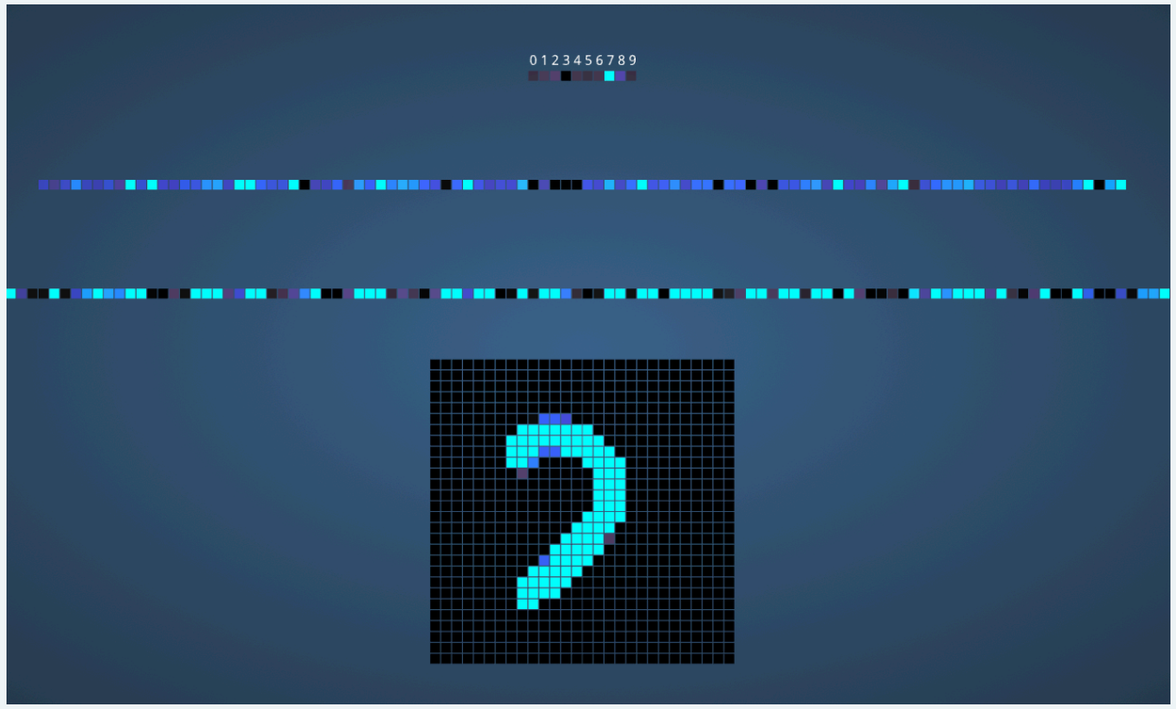
Subsymbolic representations

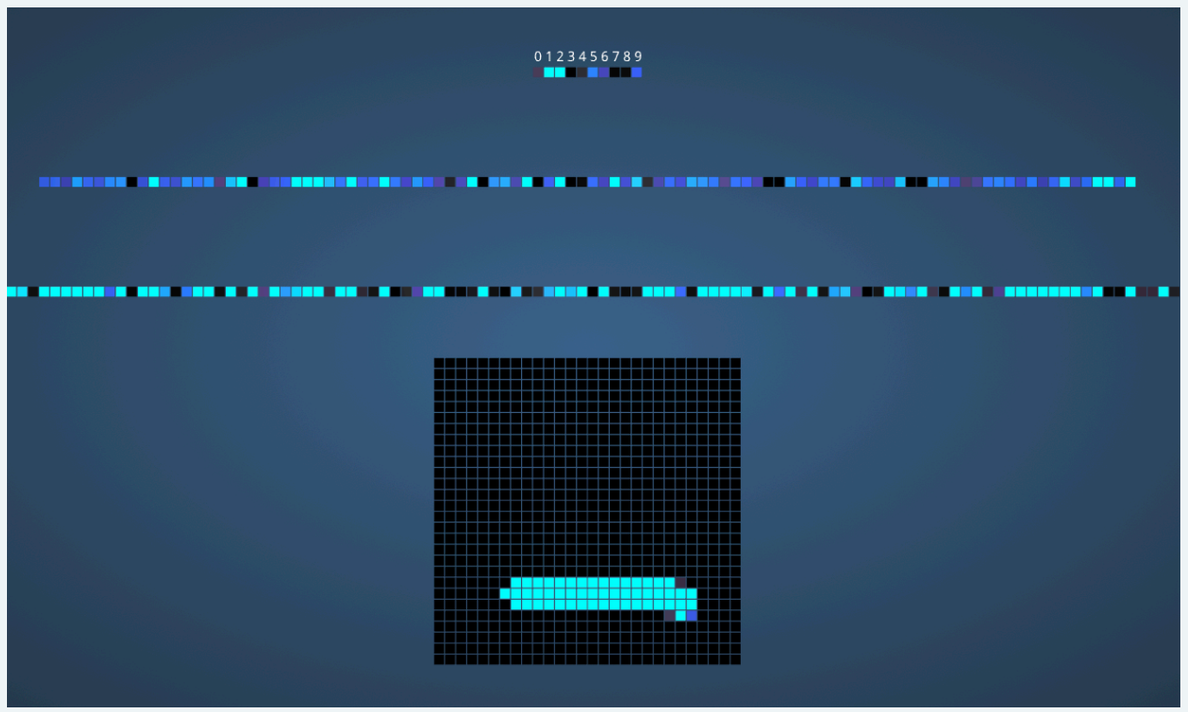
Parts of "2" light up some diff neurons, mostly similar

Not spatially organized

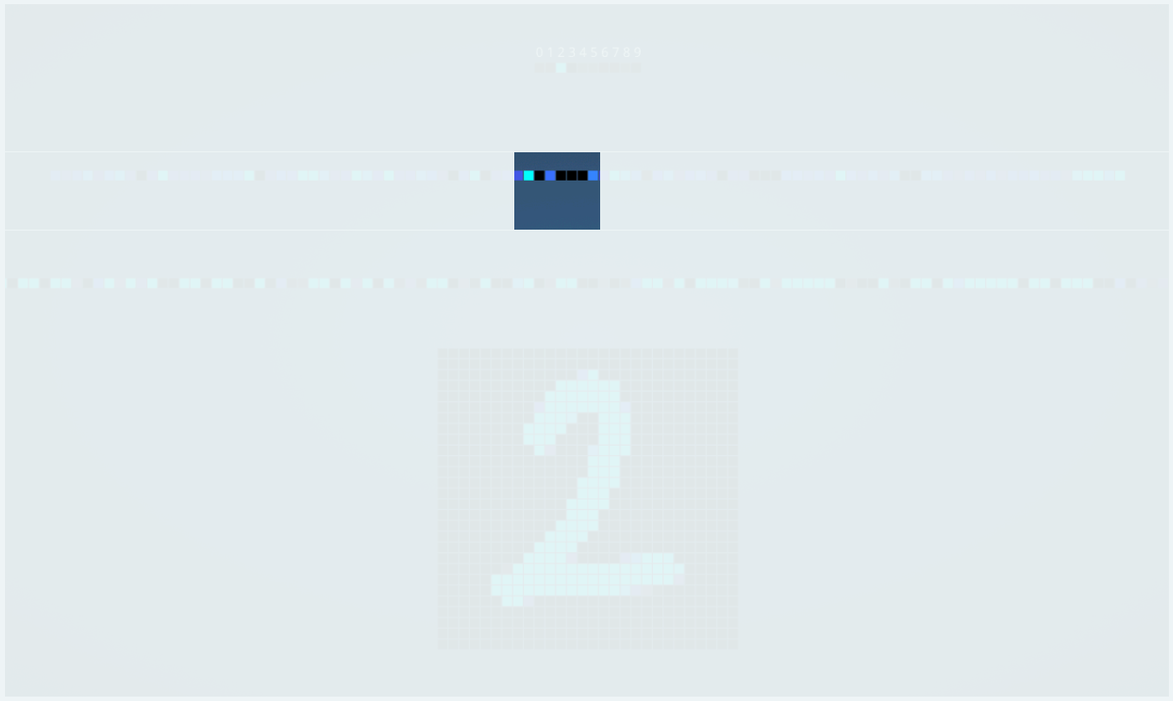
Some NNs are sequentially/ spatially dependent

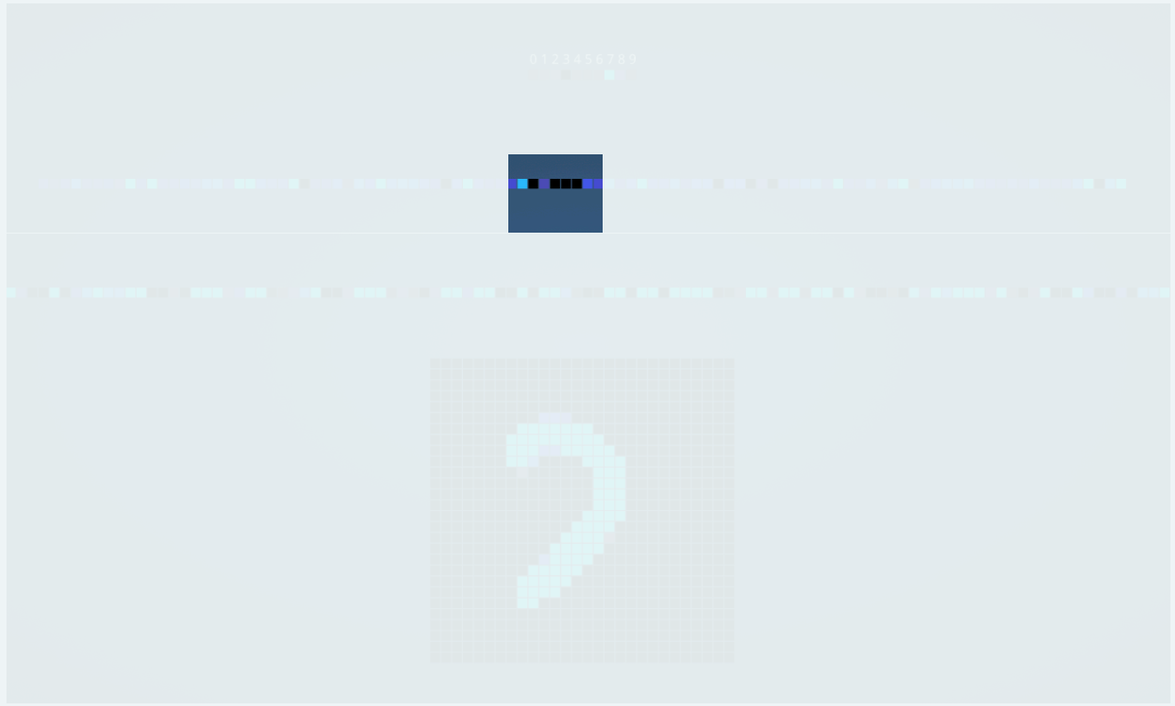


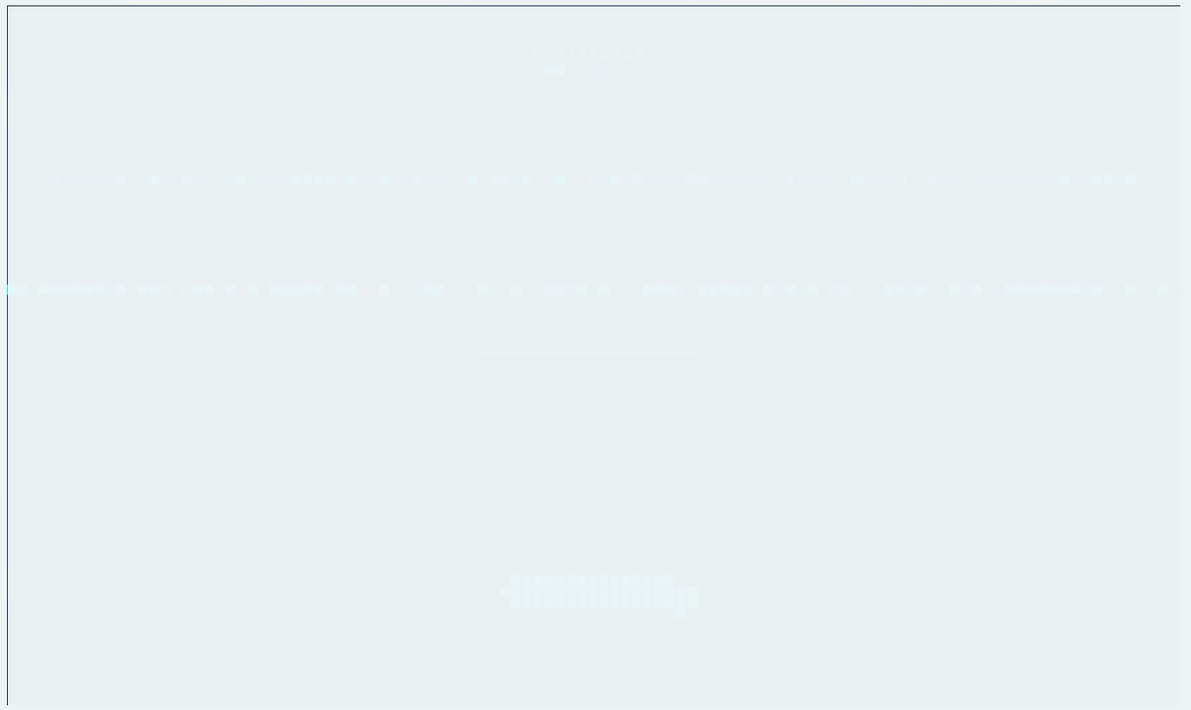




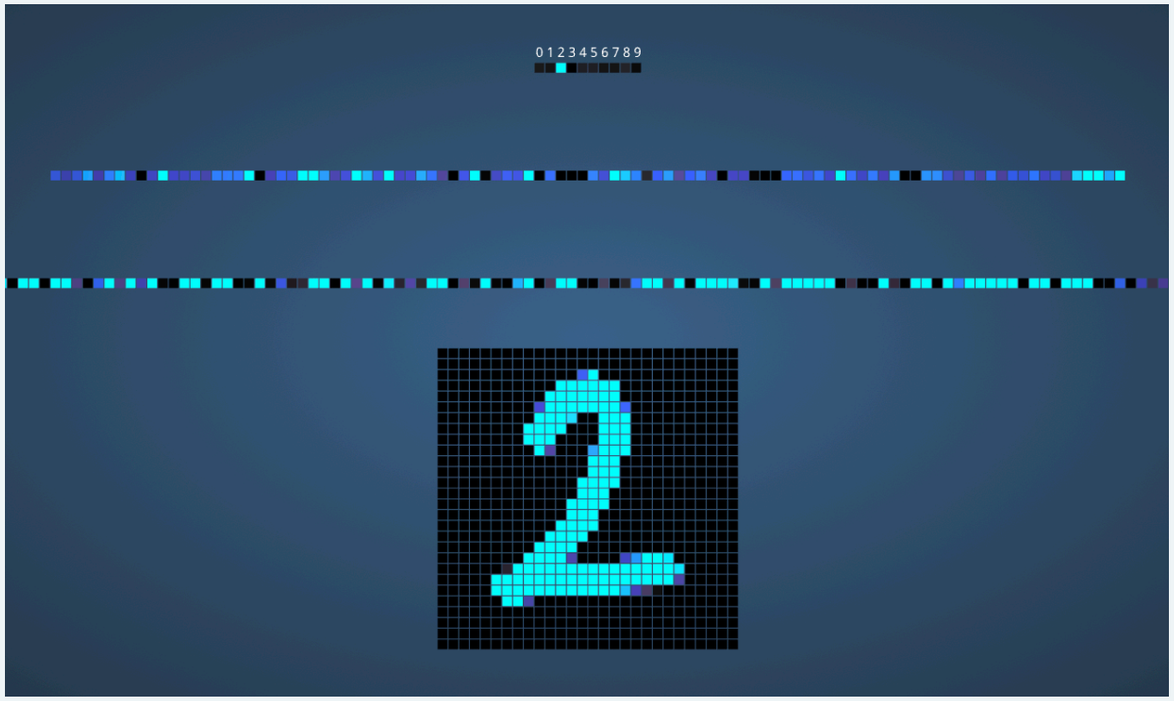
Subsymbolic representations

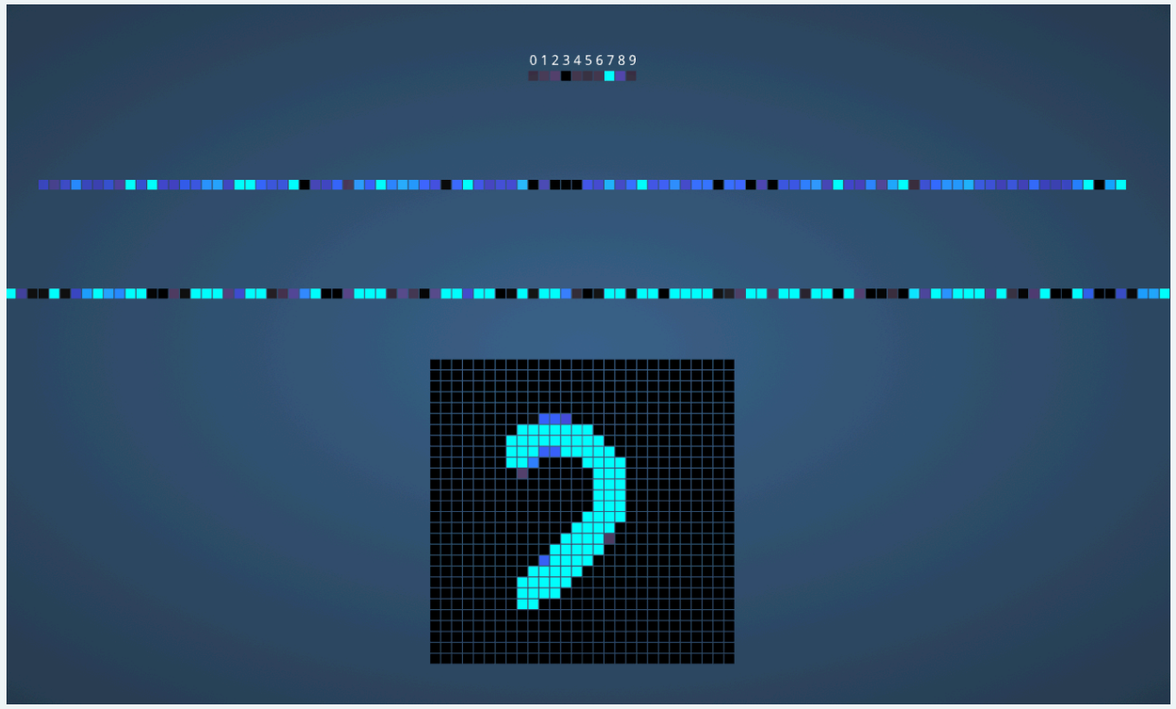


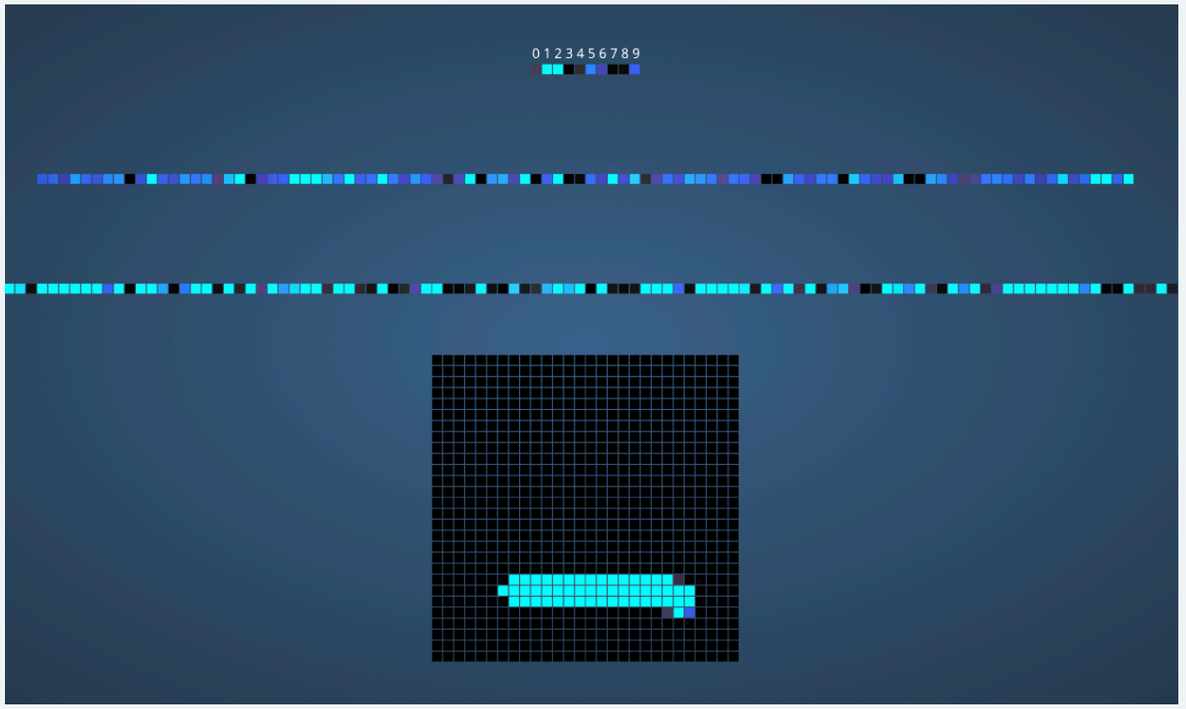




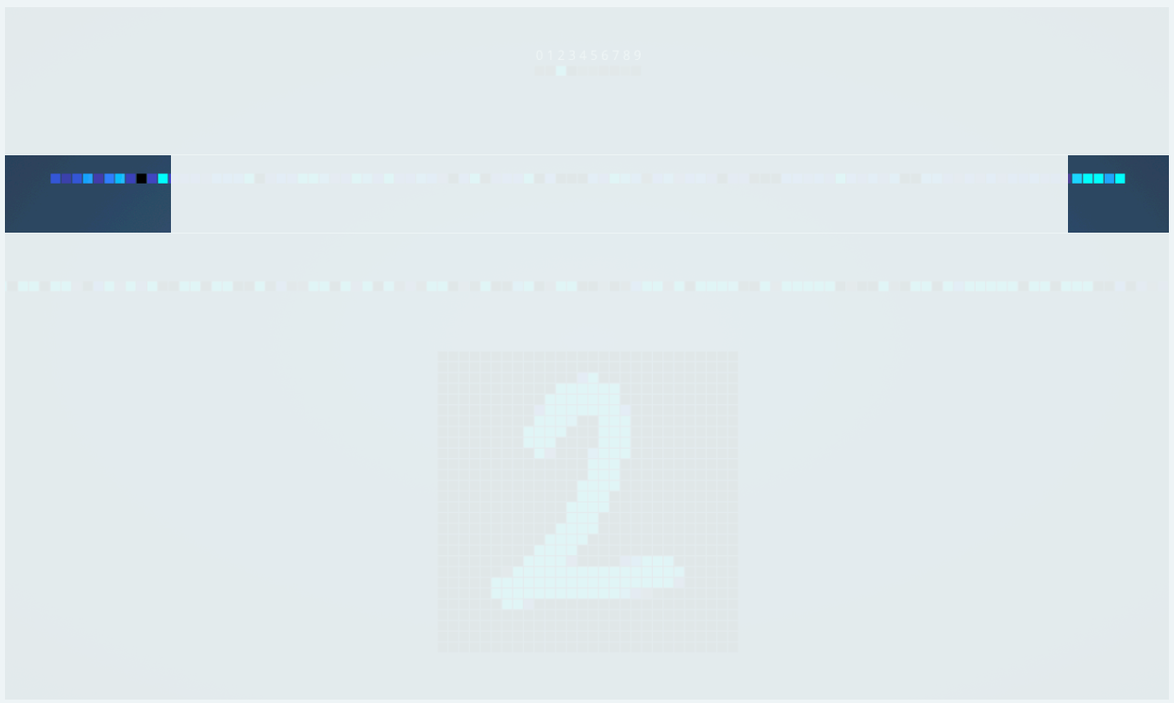
Subsymbolic representations

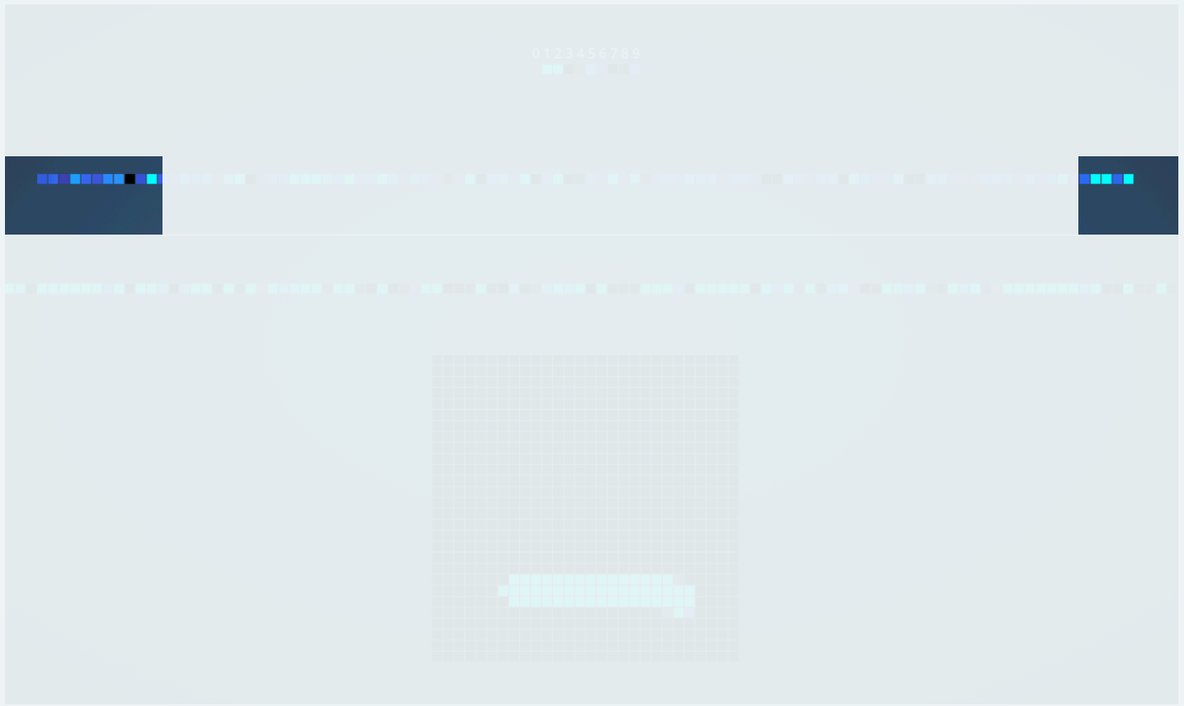




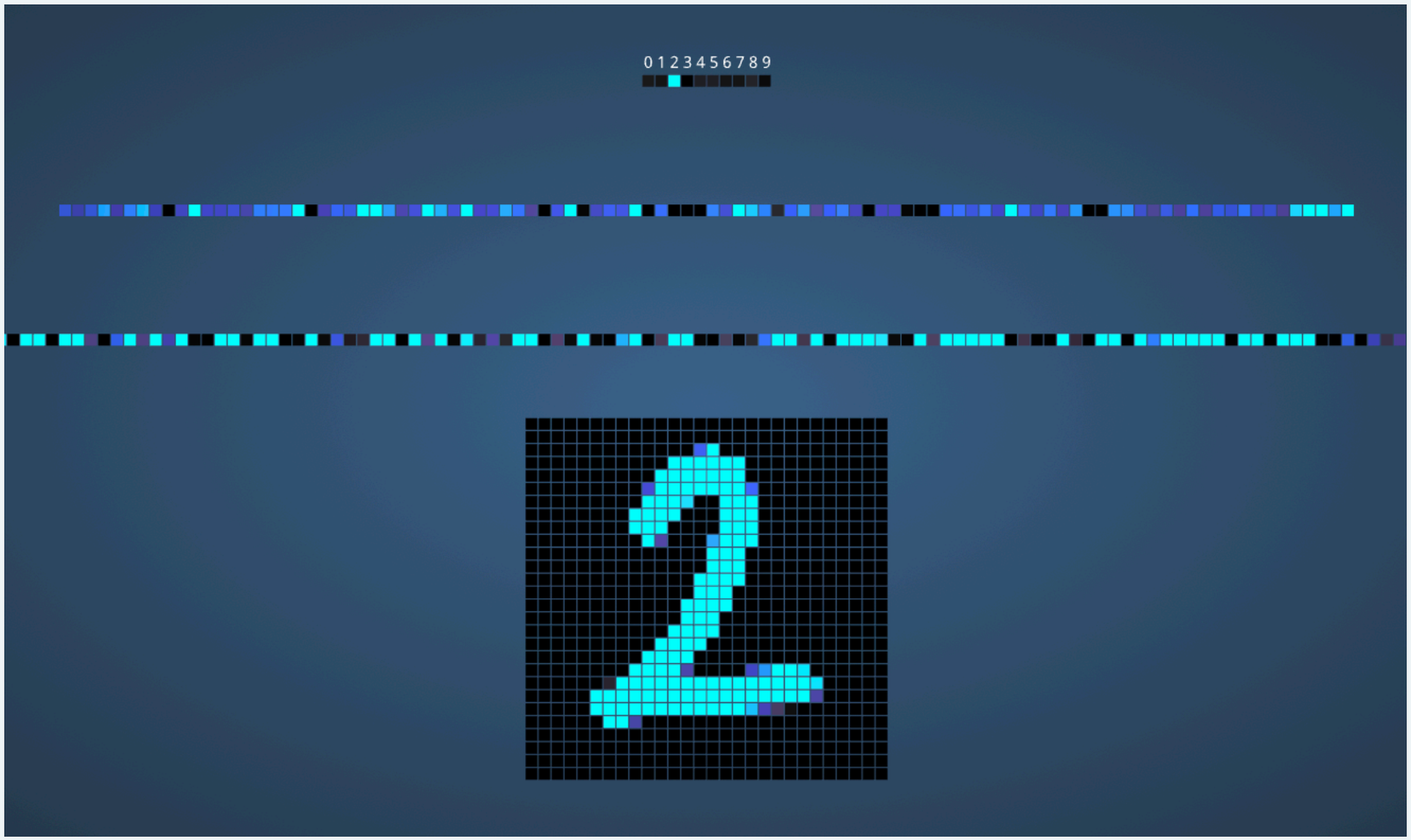


Subsymbolic representations





Graceful degradation

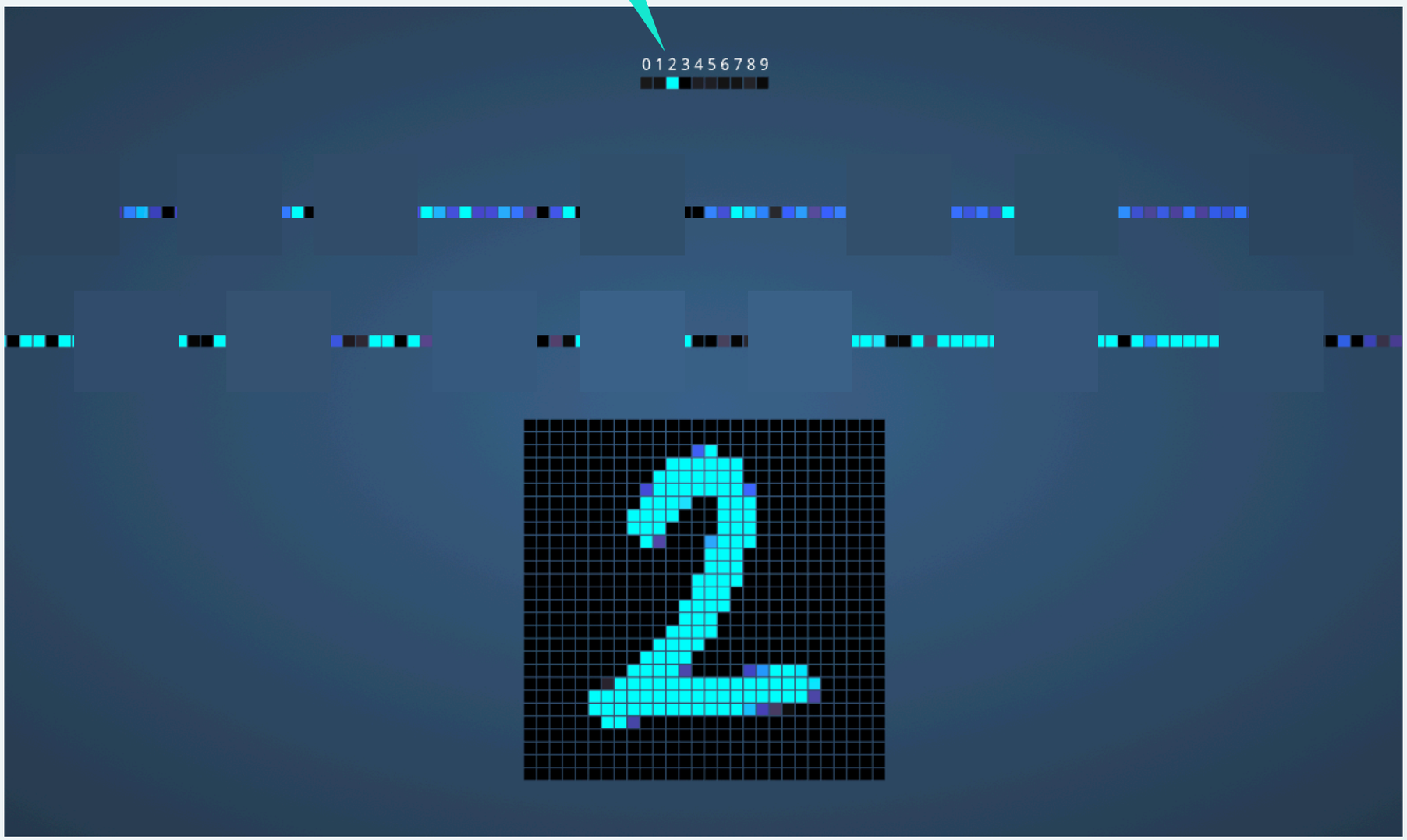


Graceful degradation

Performance doesn't plummet with removal of a few units

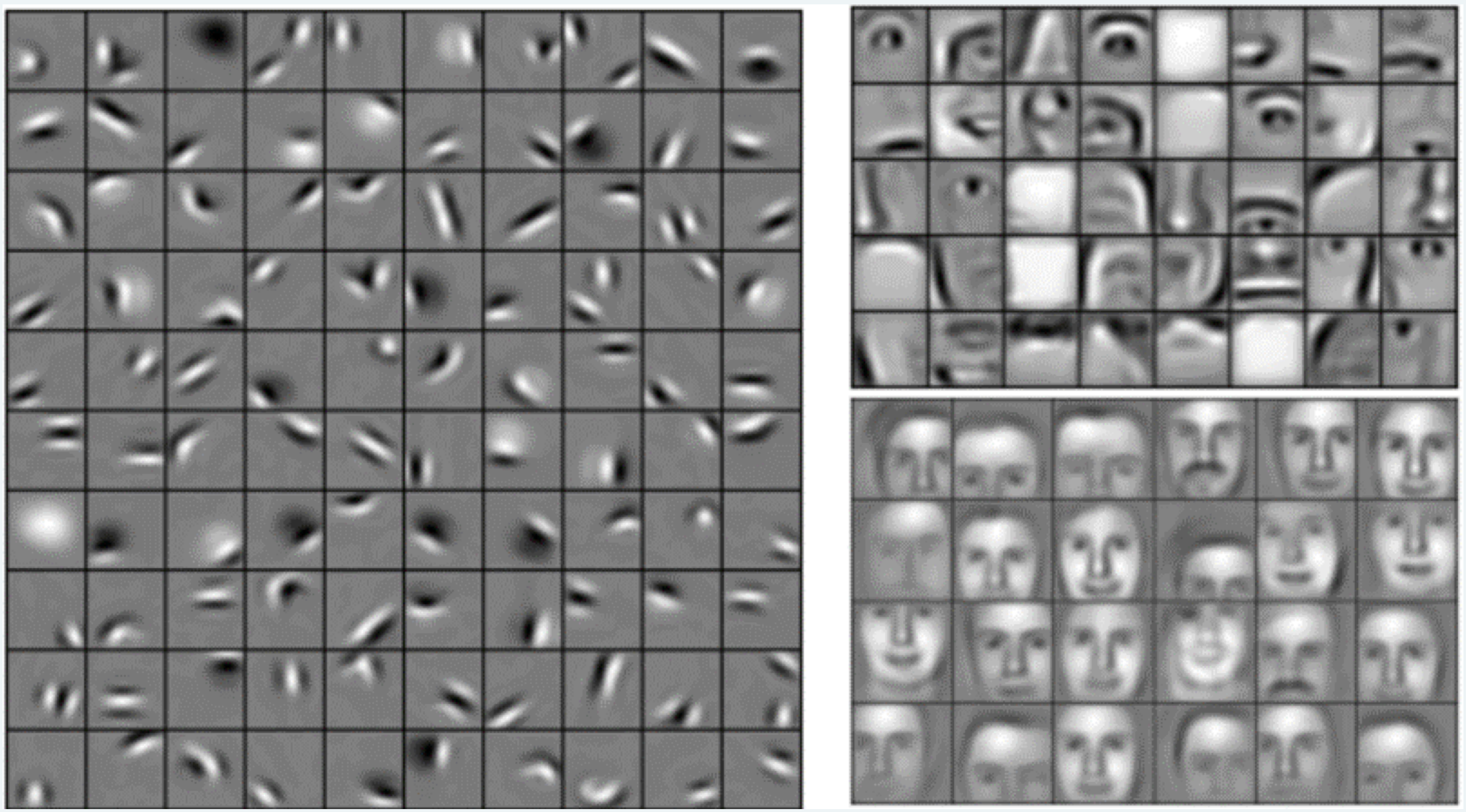
I’m still OK!

Not the case with PSS



Connectionist networks

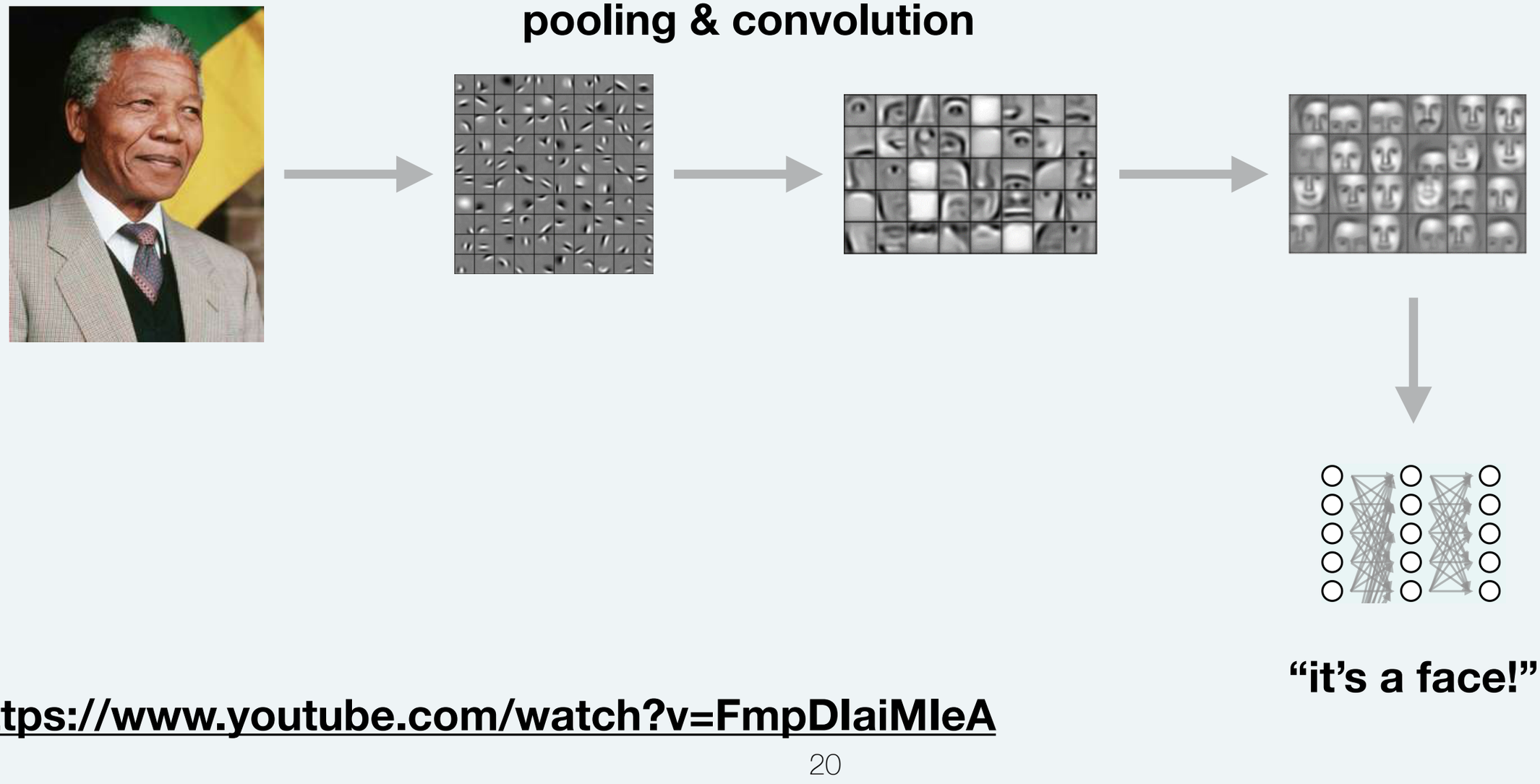
Convolutional Network



Connectionist networks

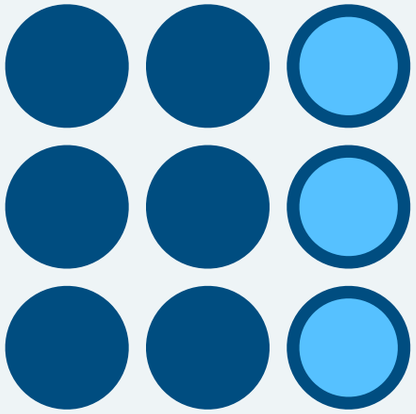
Convolutional Network

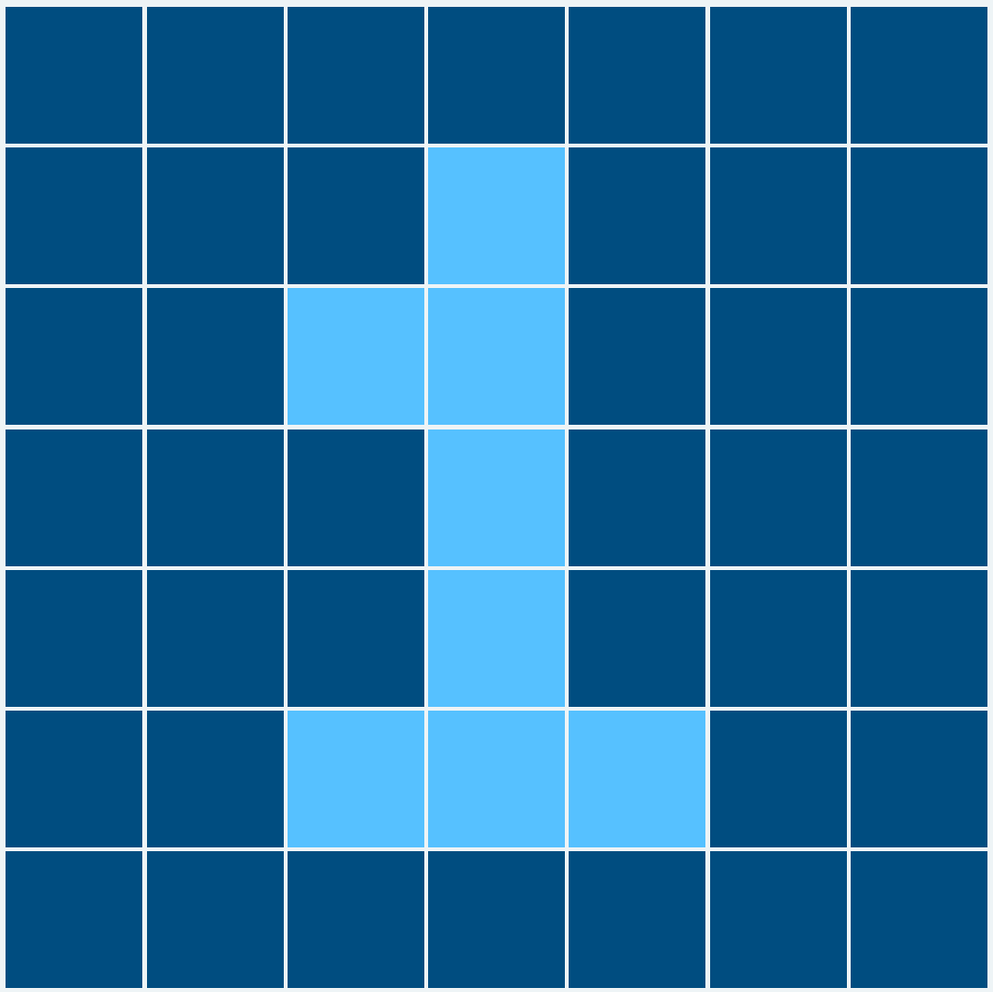
convolution **pooling & convolution** **pooling & convolution**



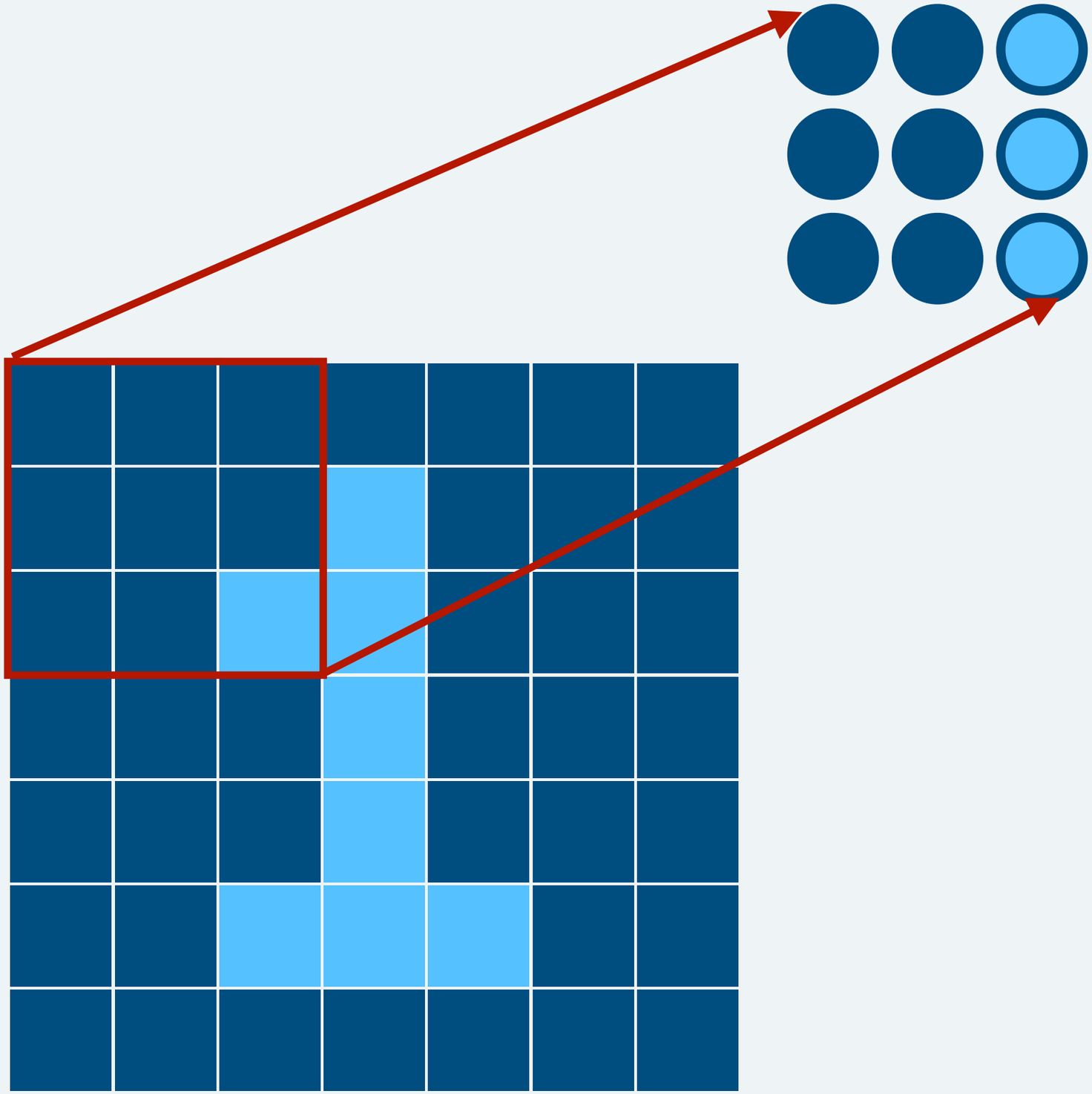
https://www.youtube.com/watch?v=FmpDIaiMIeA

Convolution

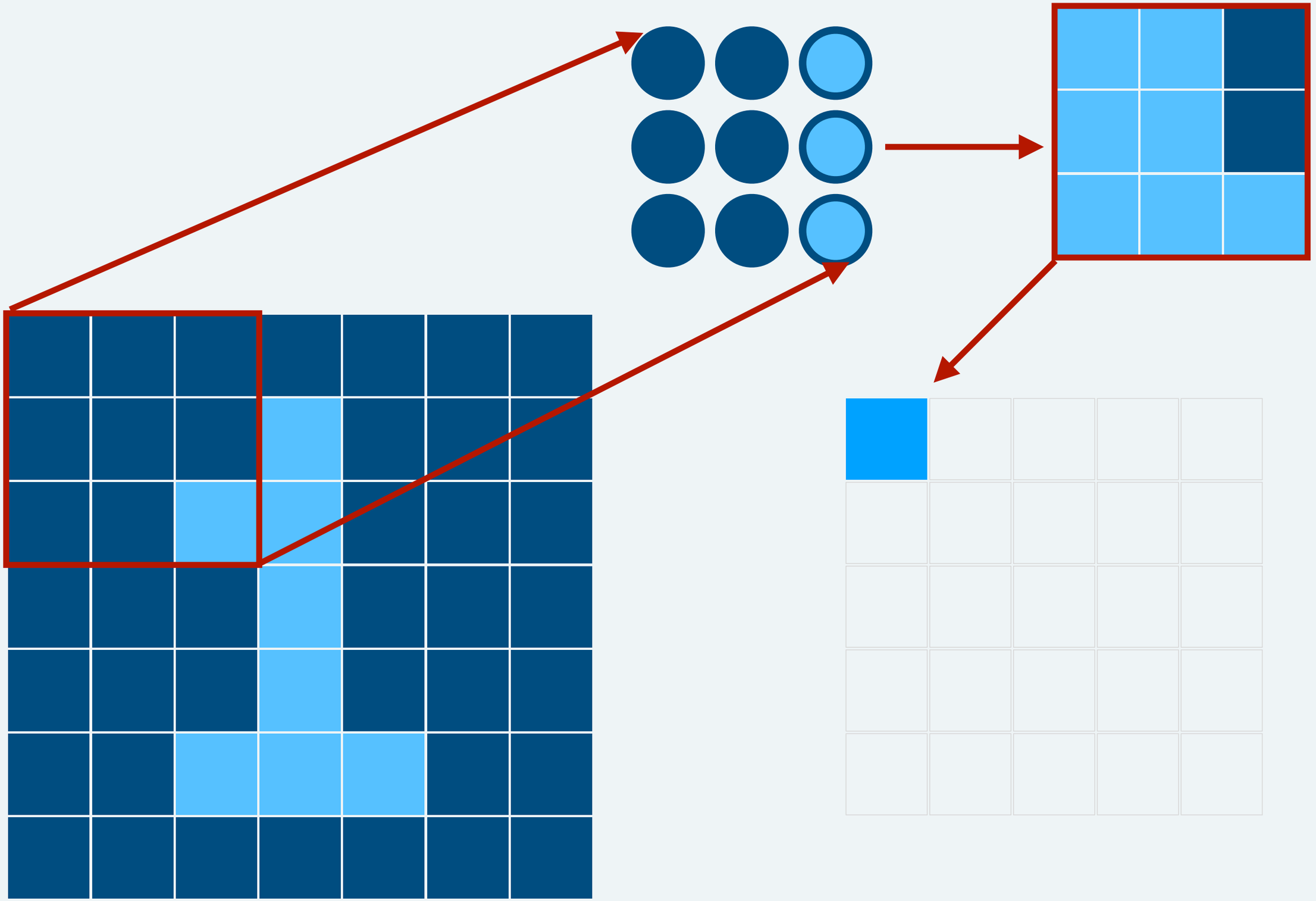




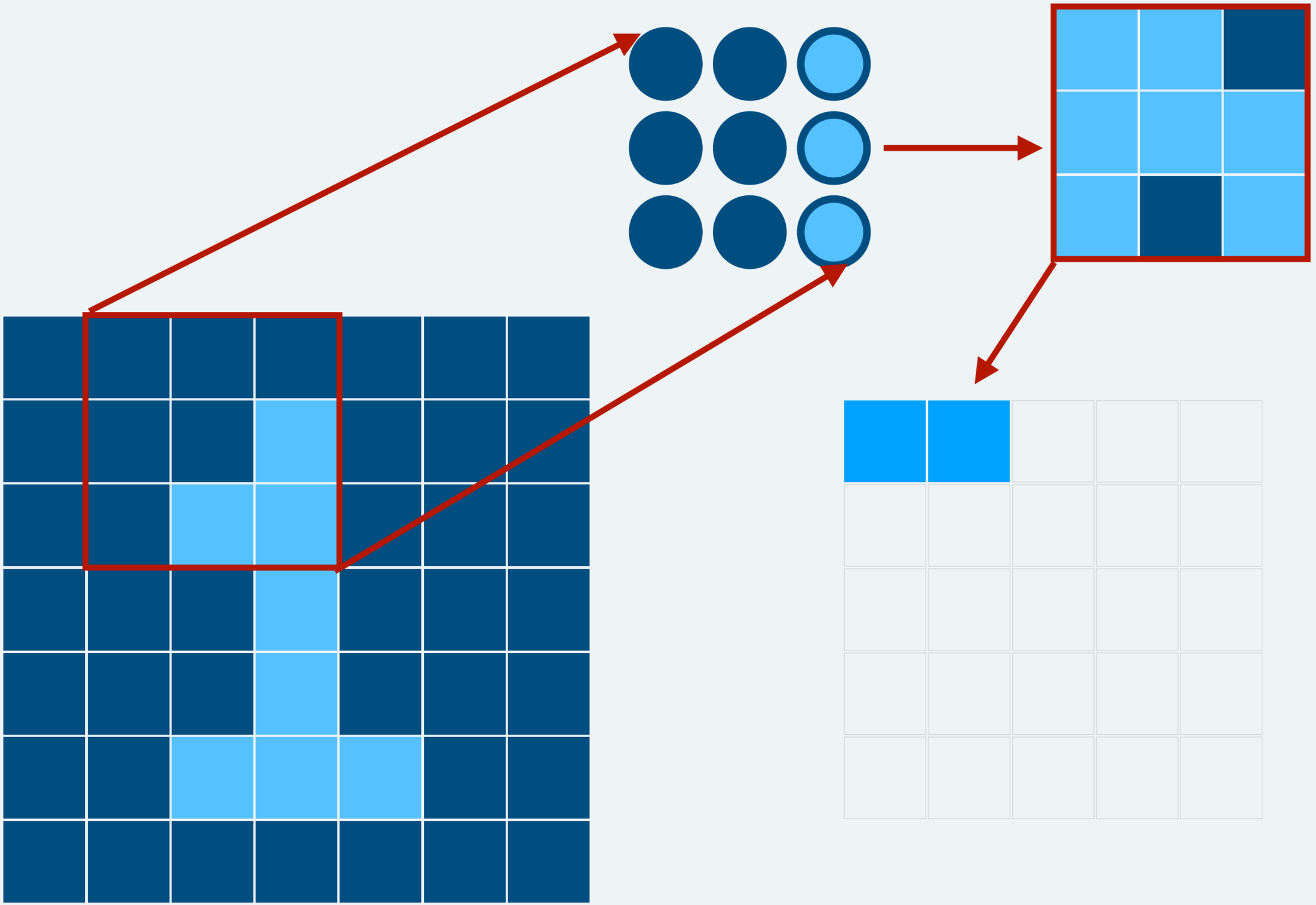
Convolution



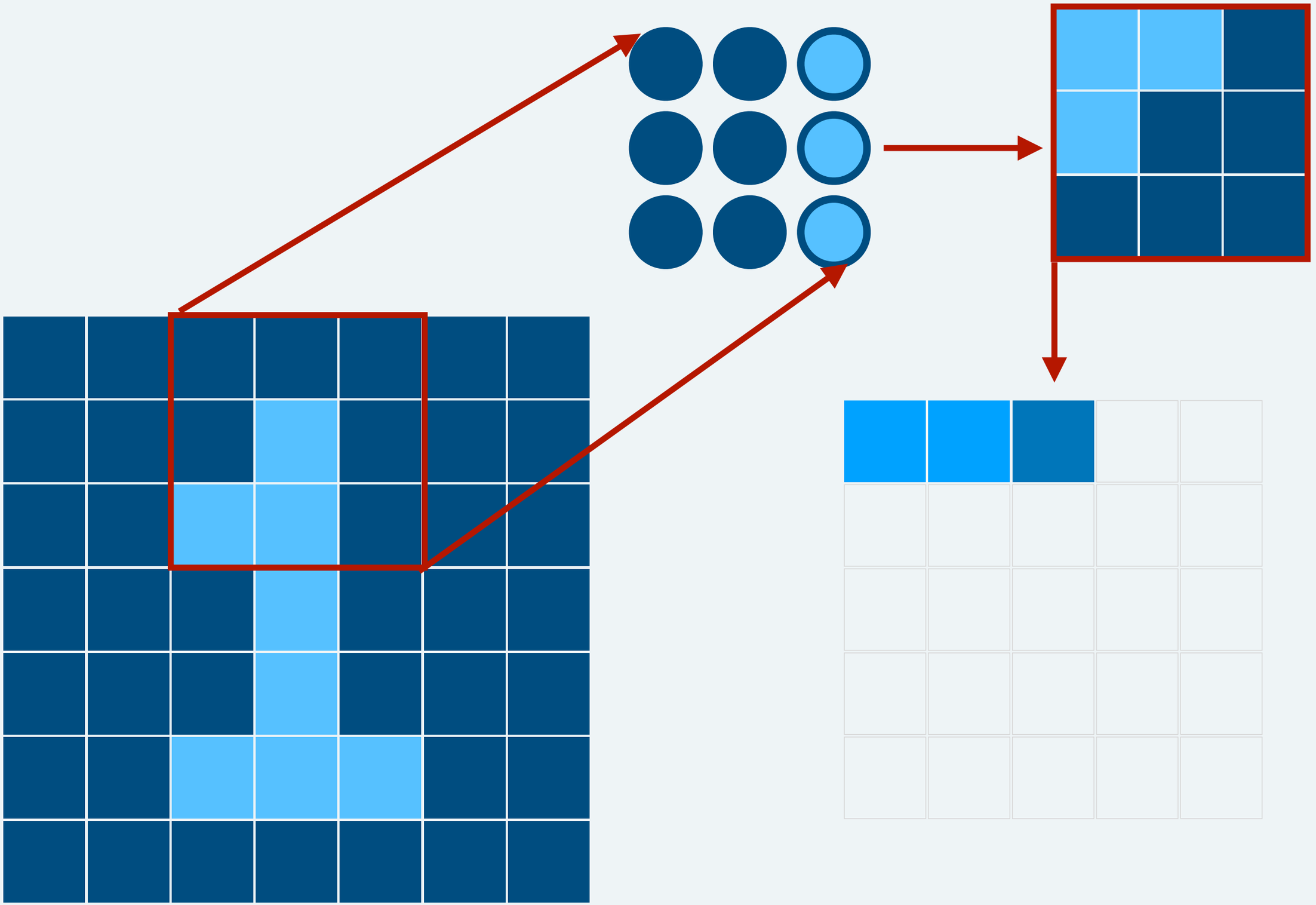
Convolution



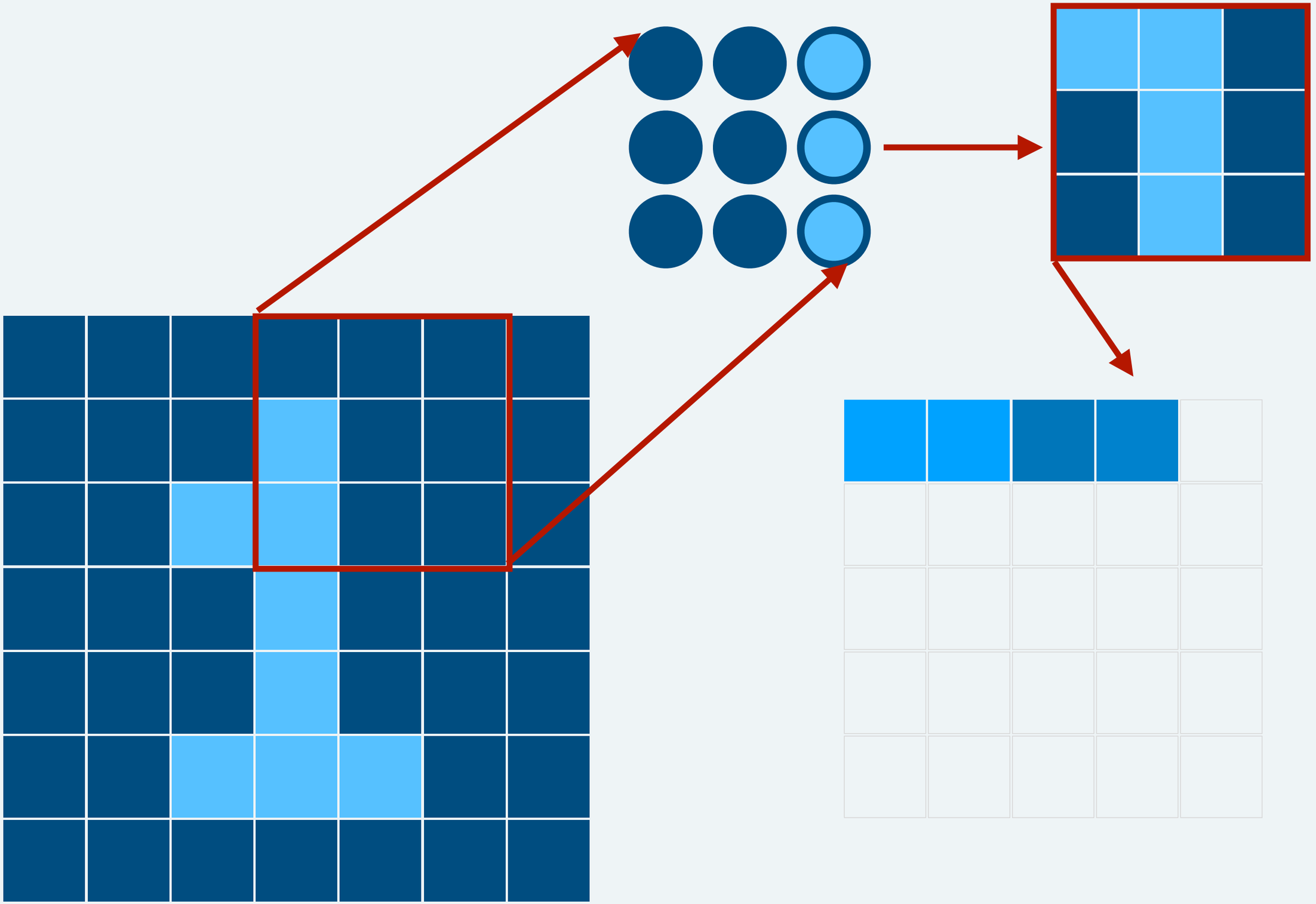
Convolution



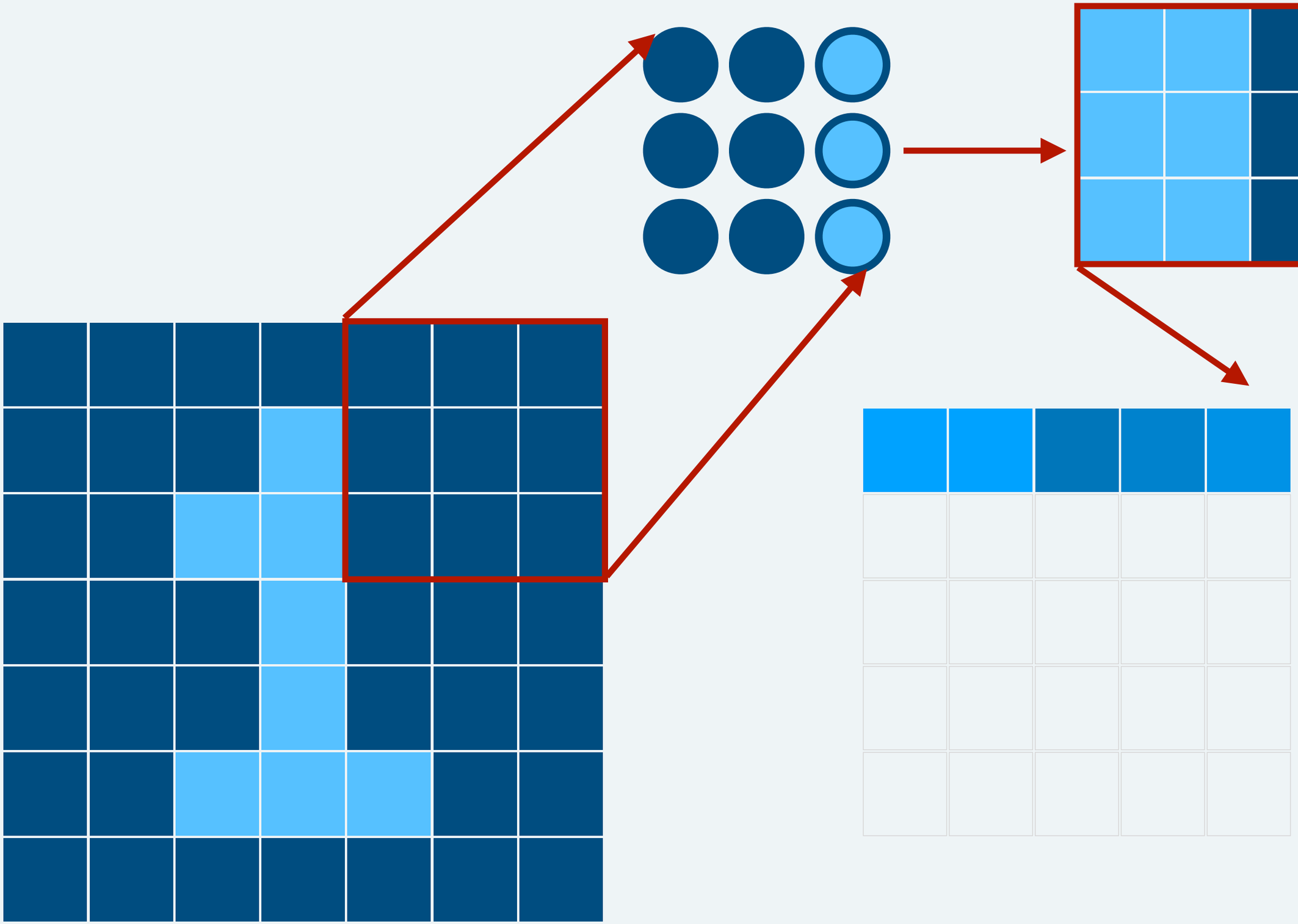
Convolution



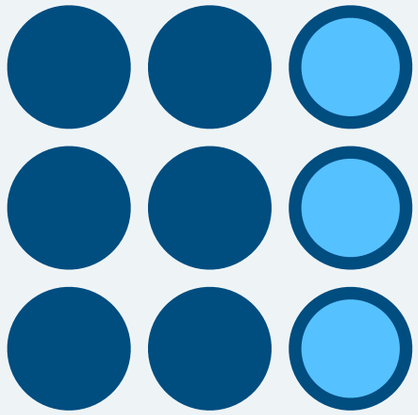
Convolution

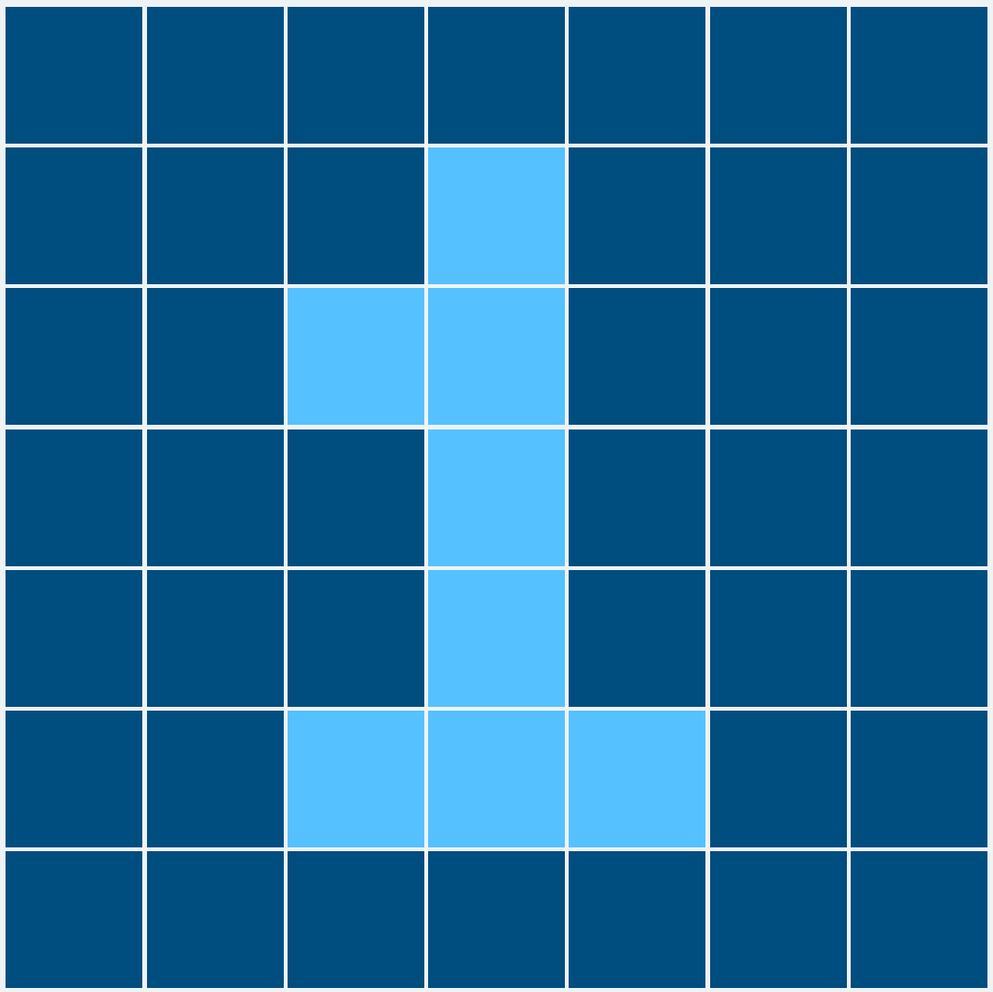


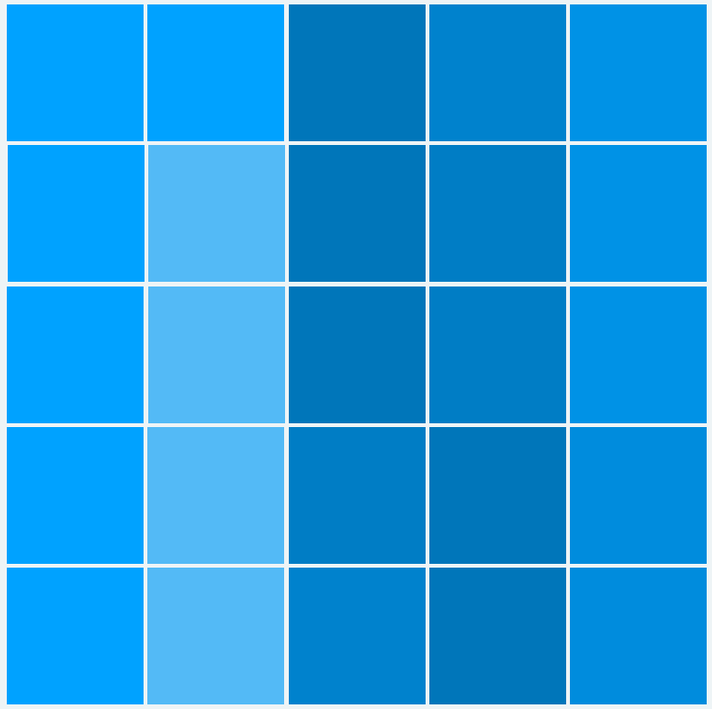
Convolution



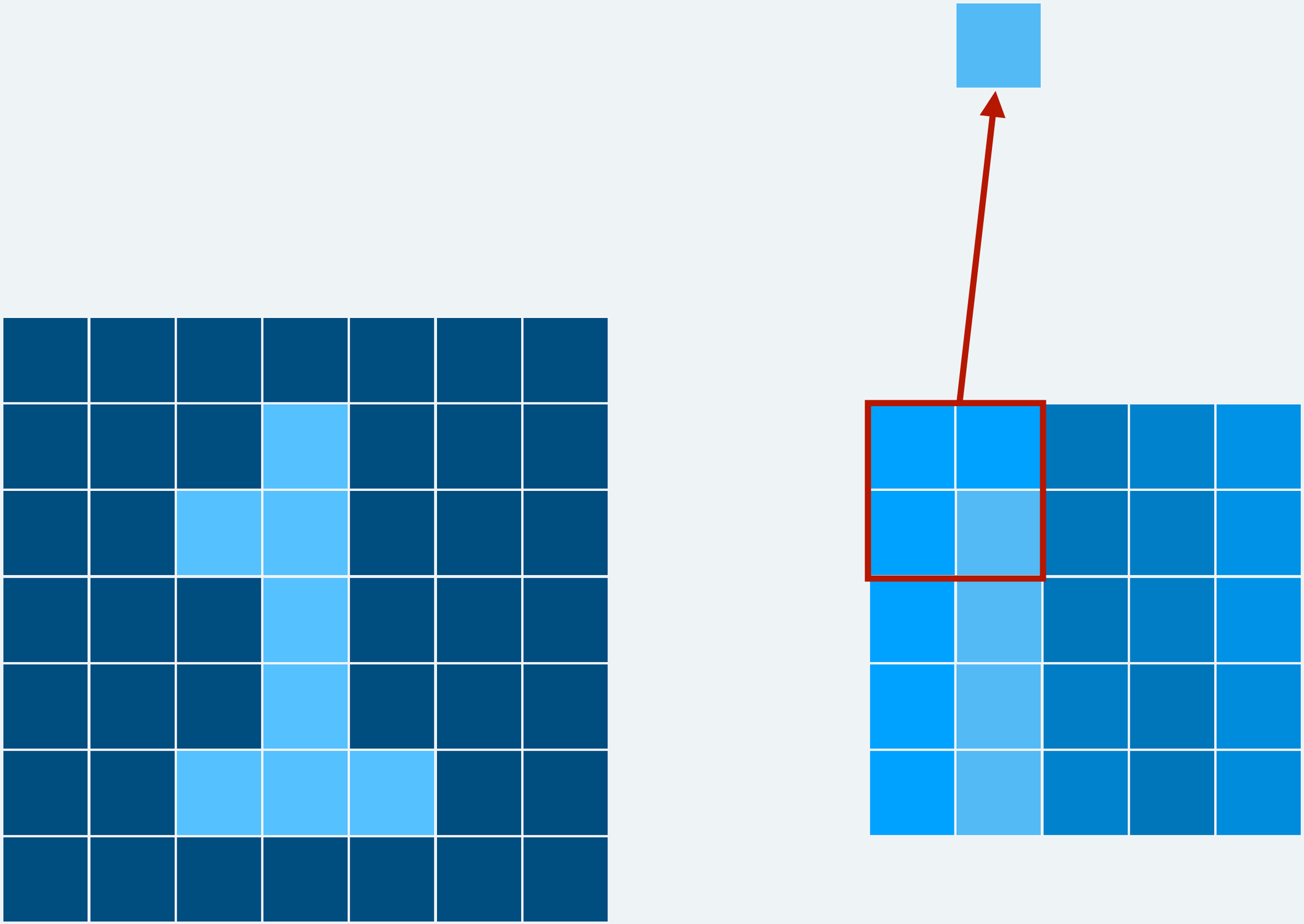
Convolution



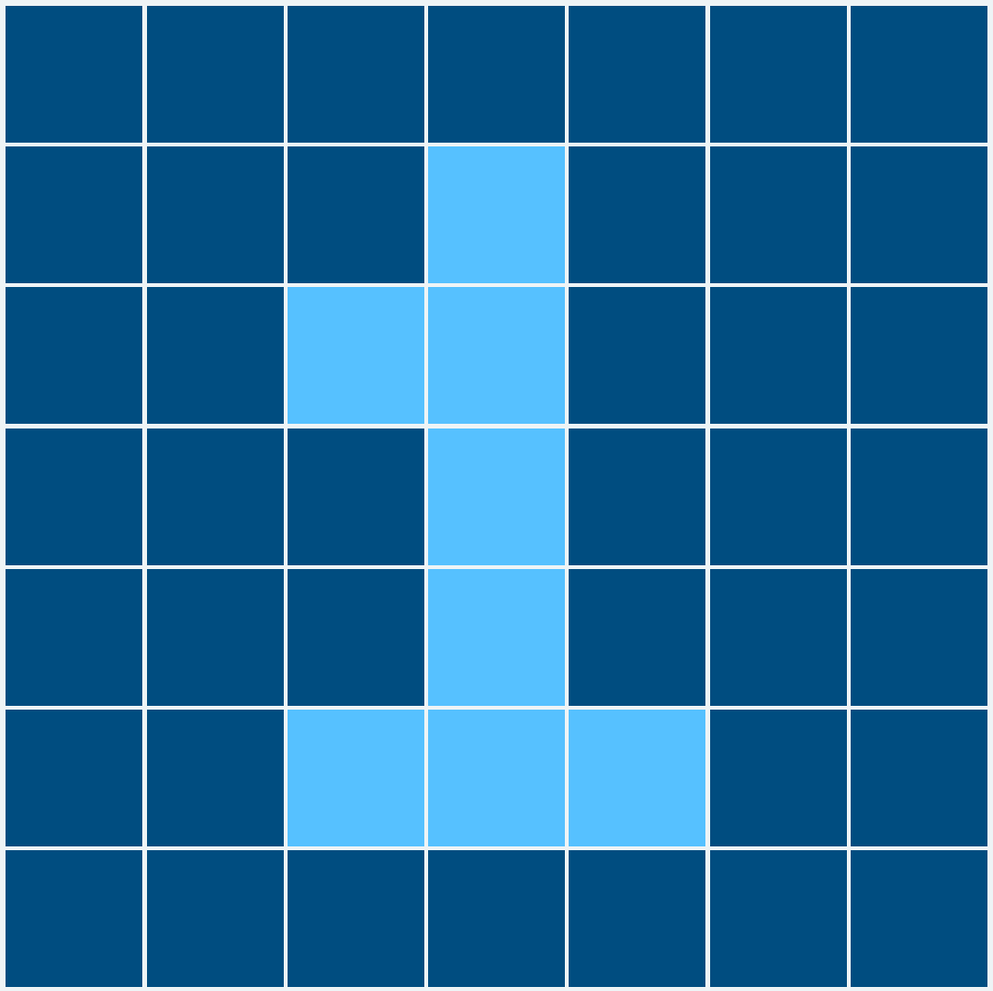


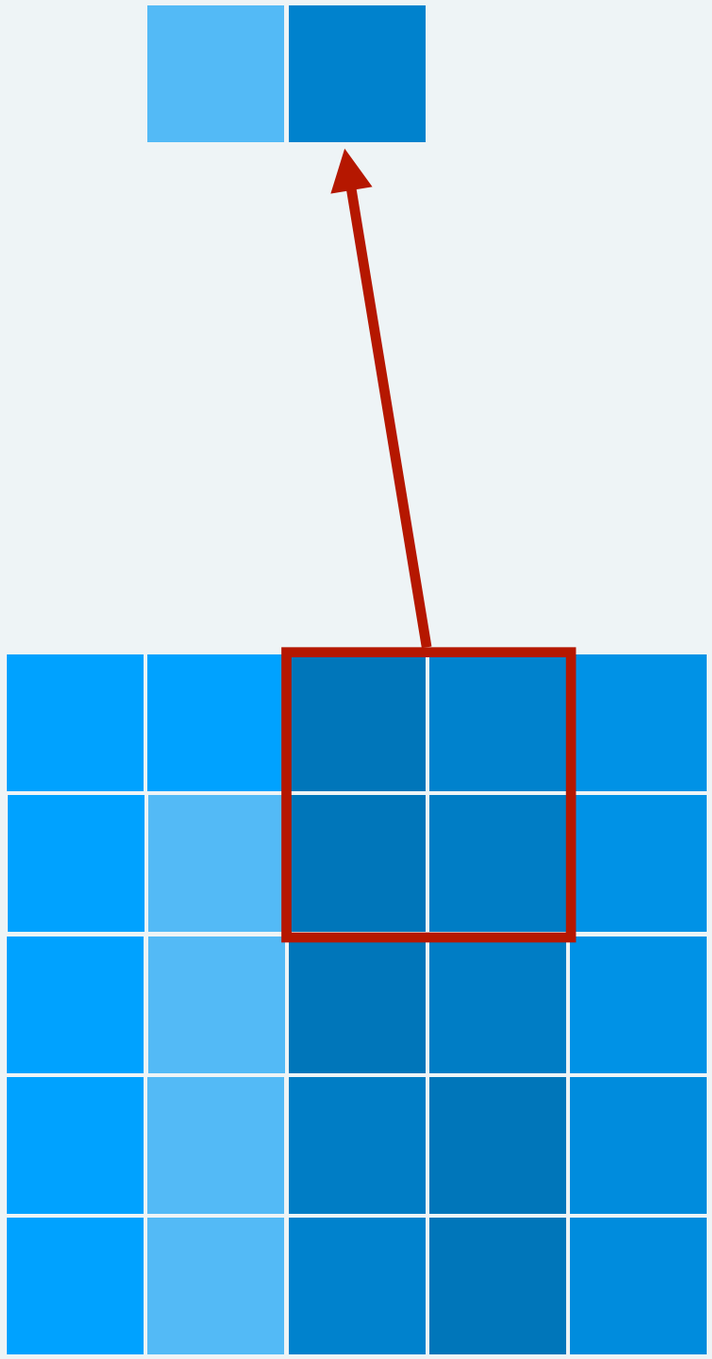


Pooling

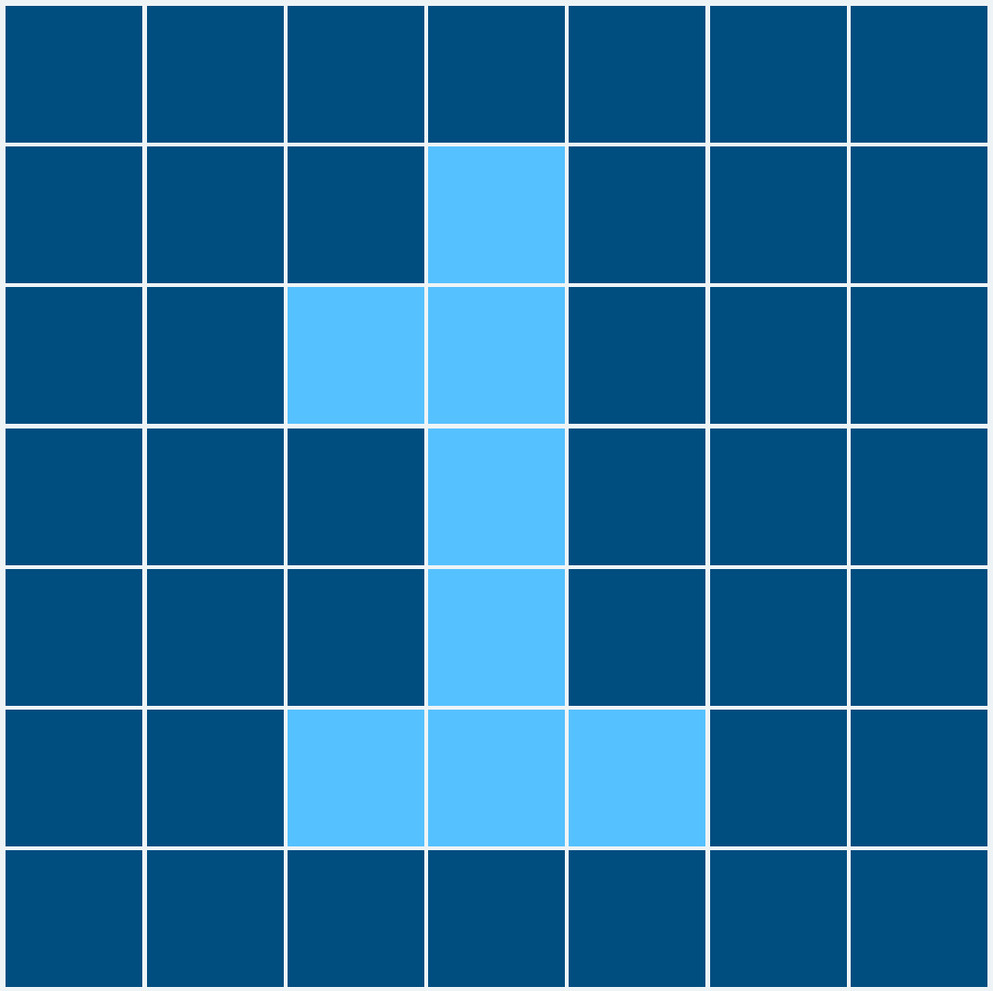


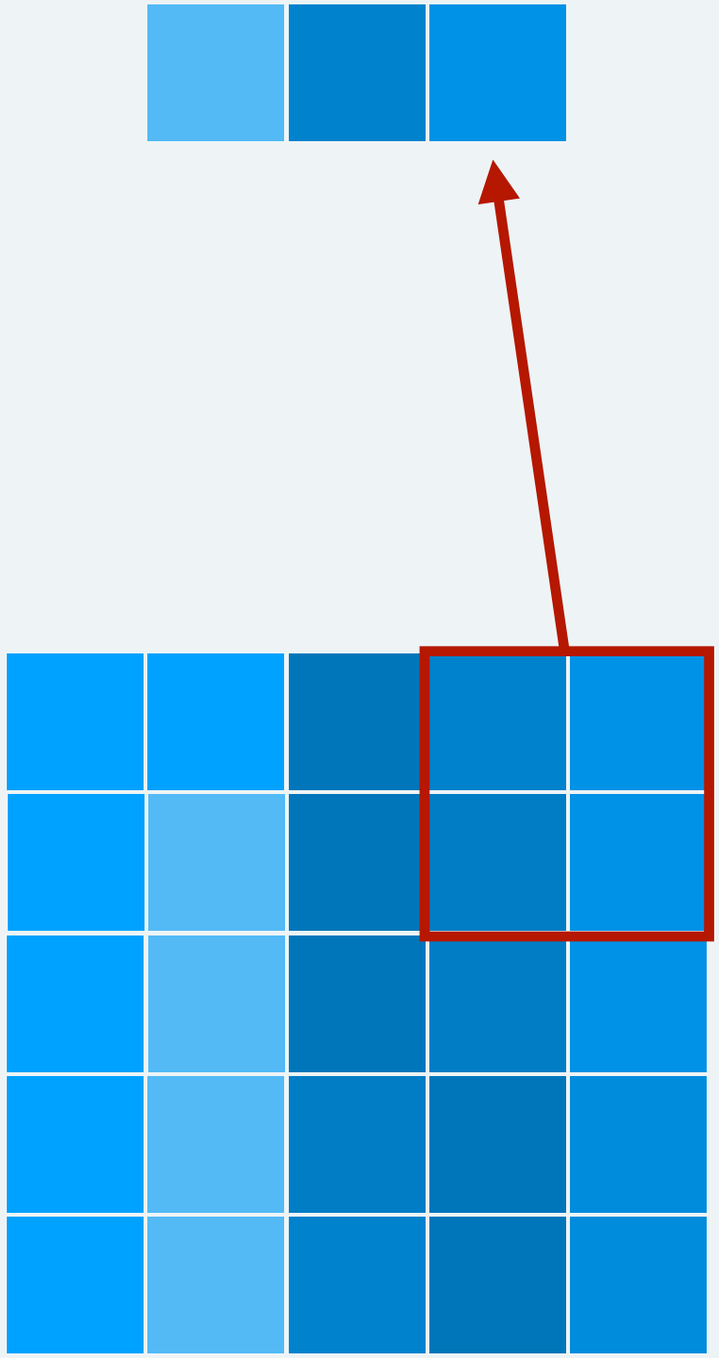
Pooling



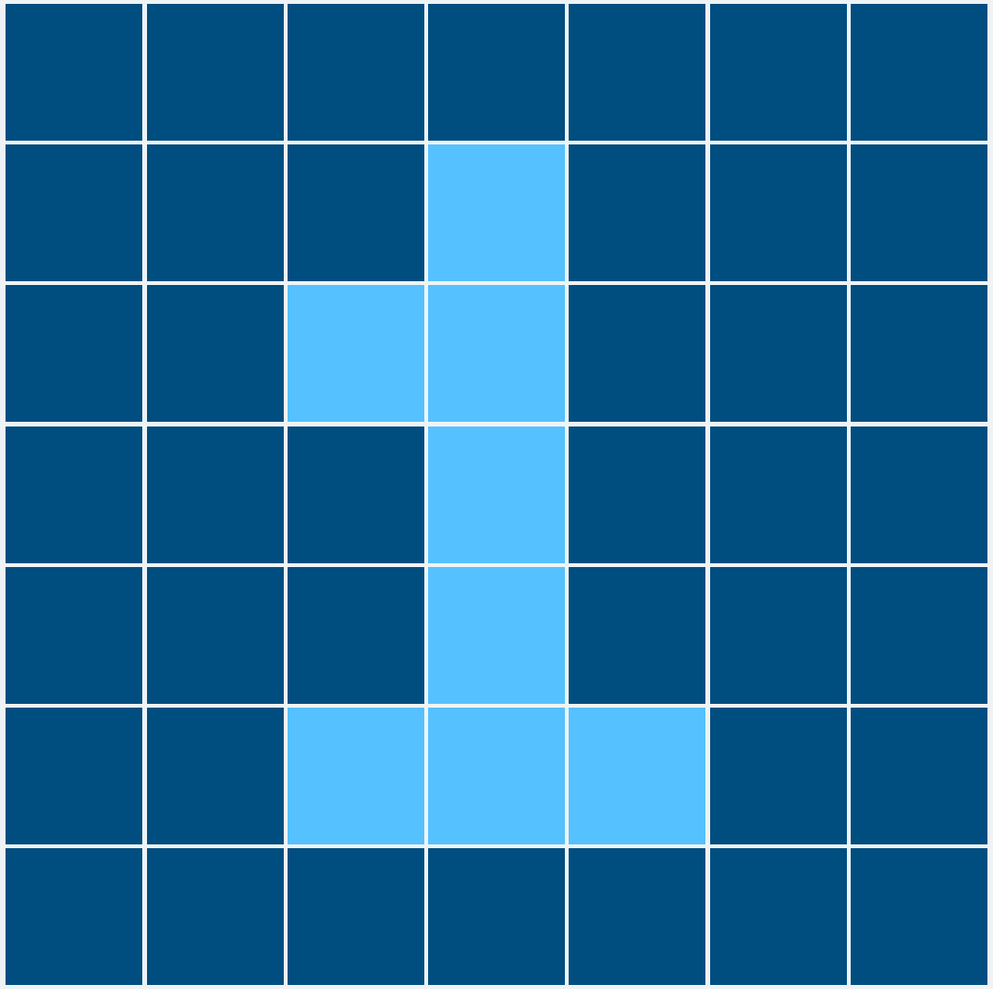


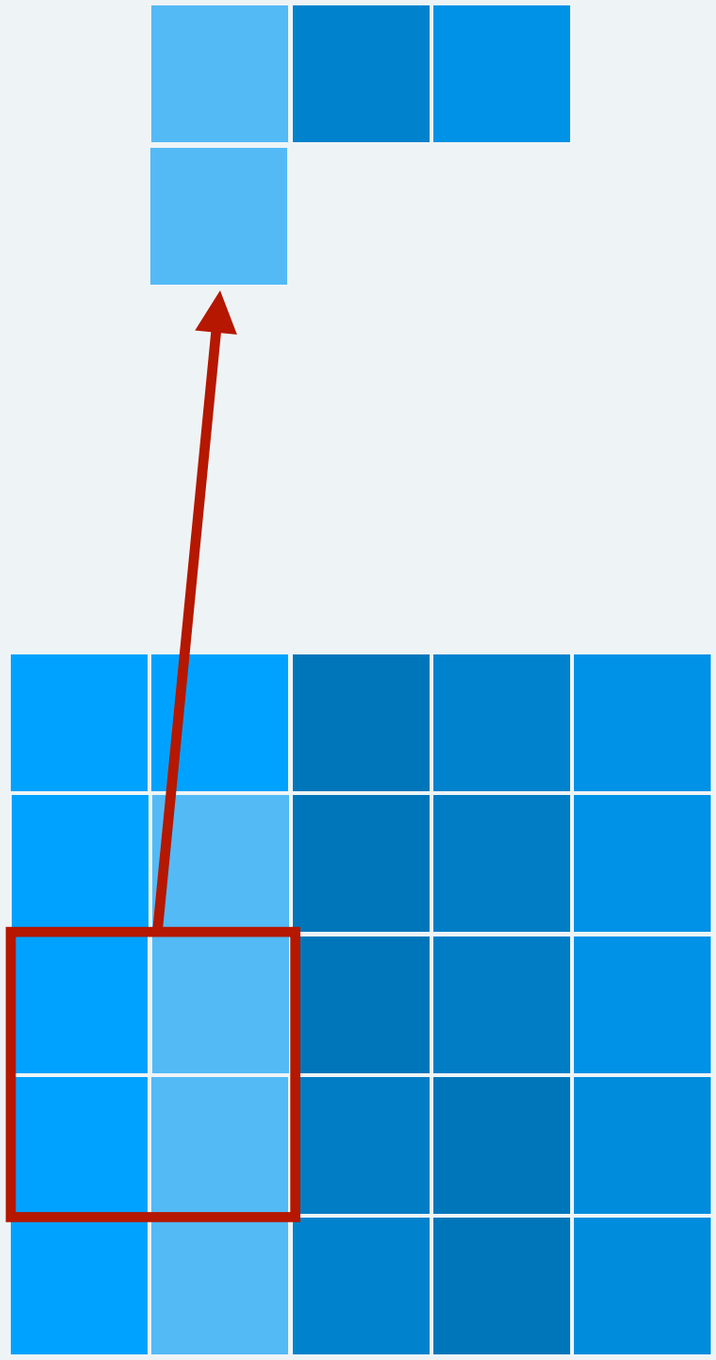
Pooling



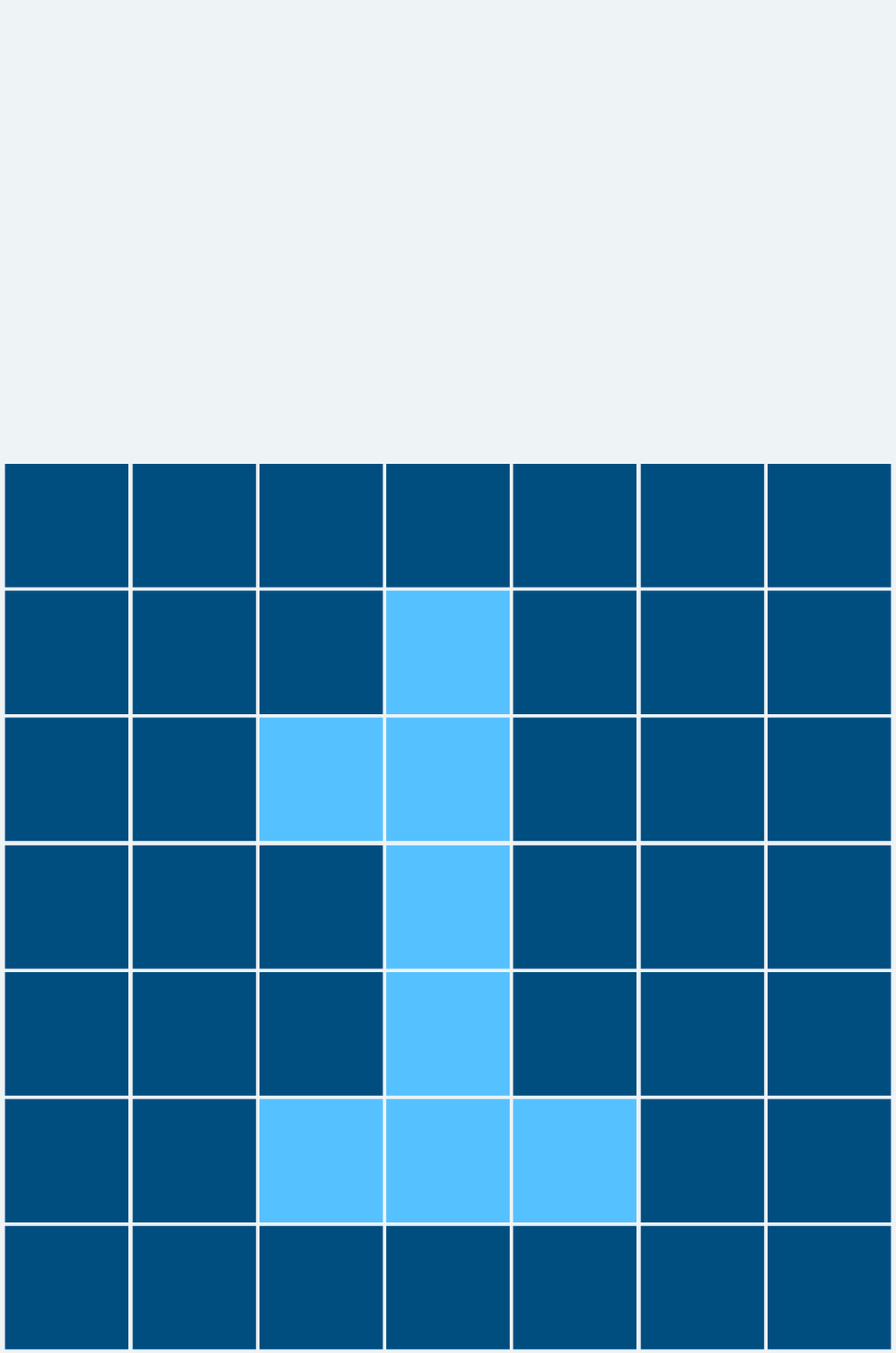


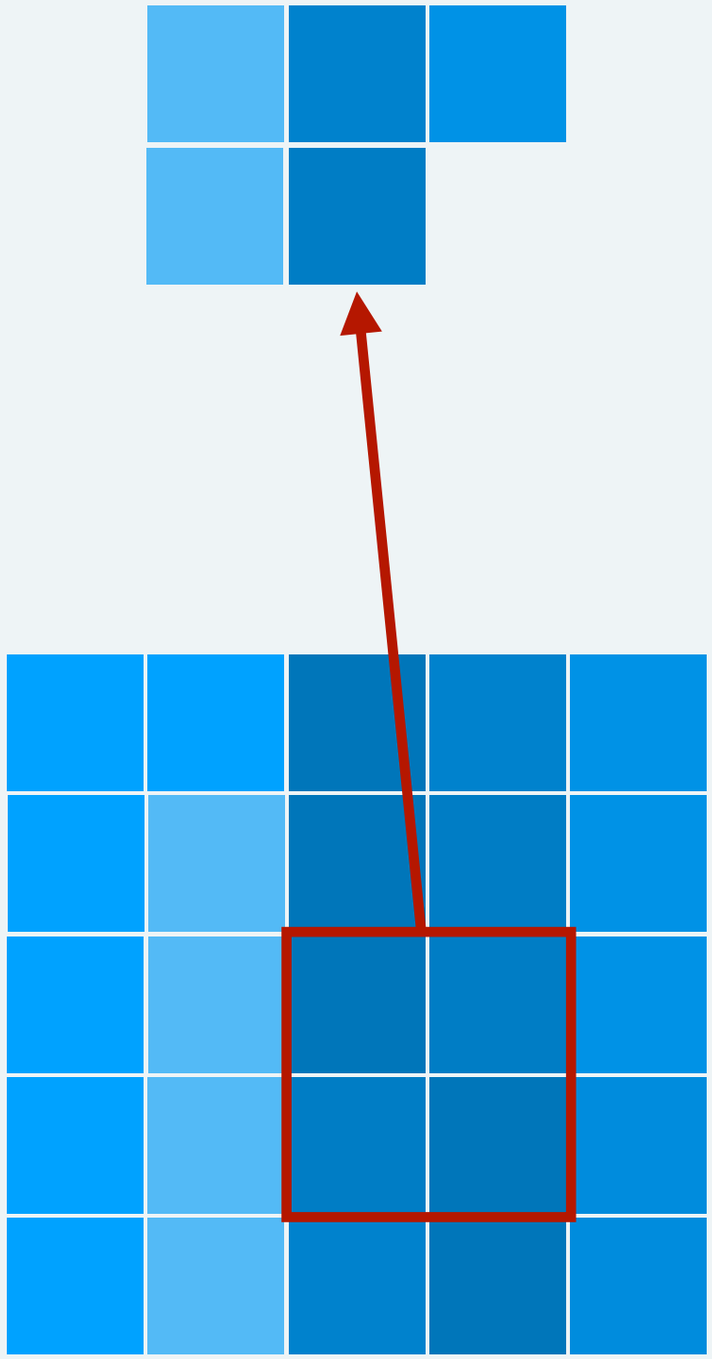
Pooling



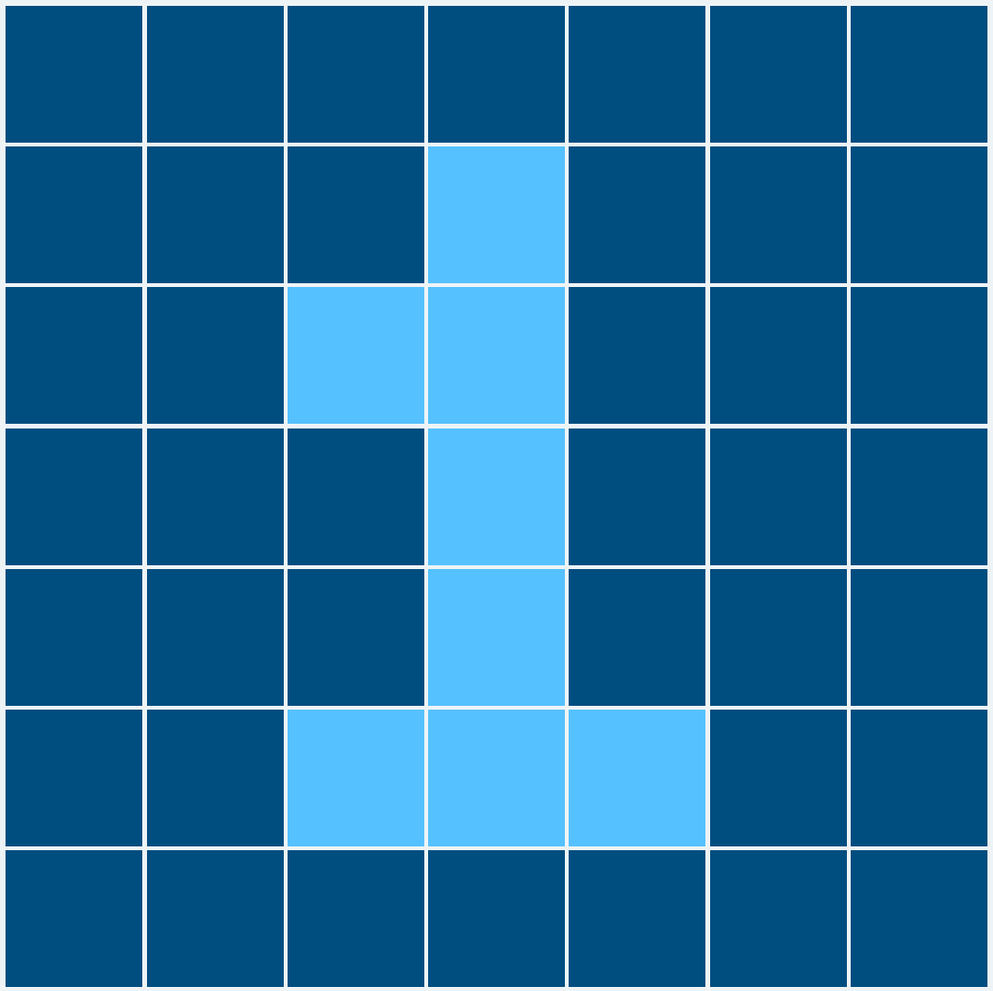


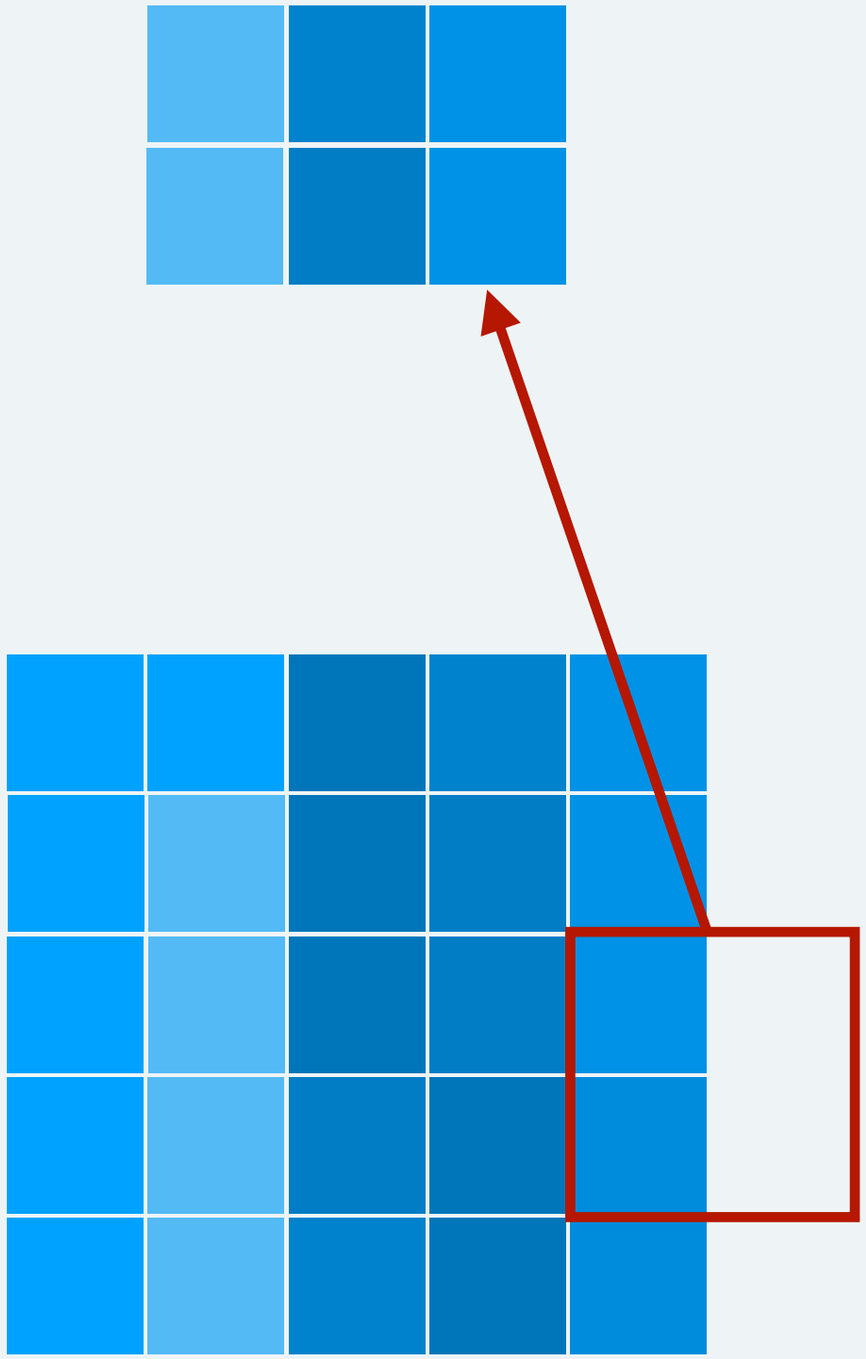
Pooling

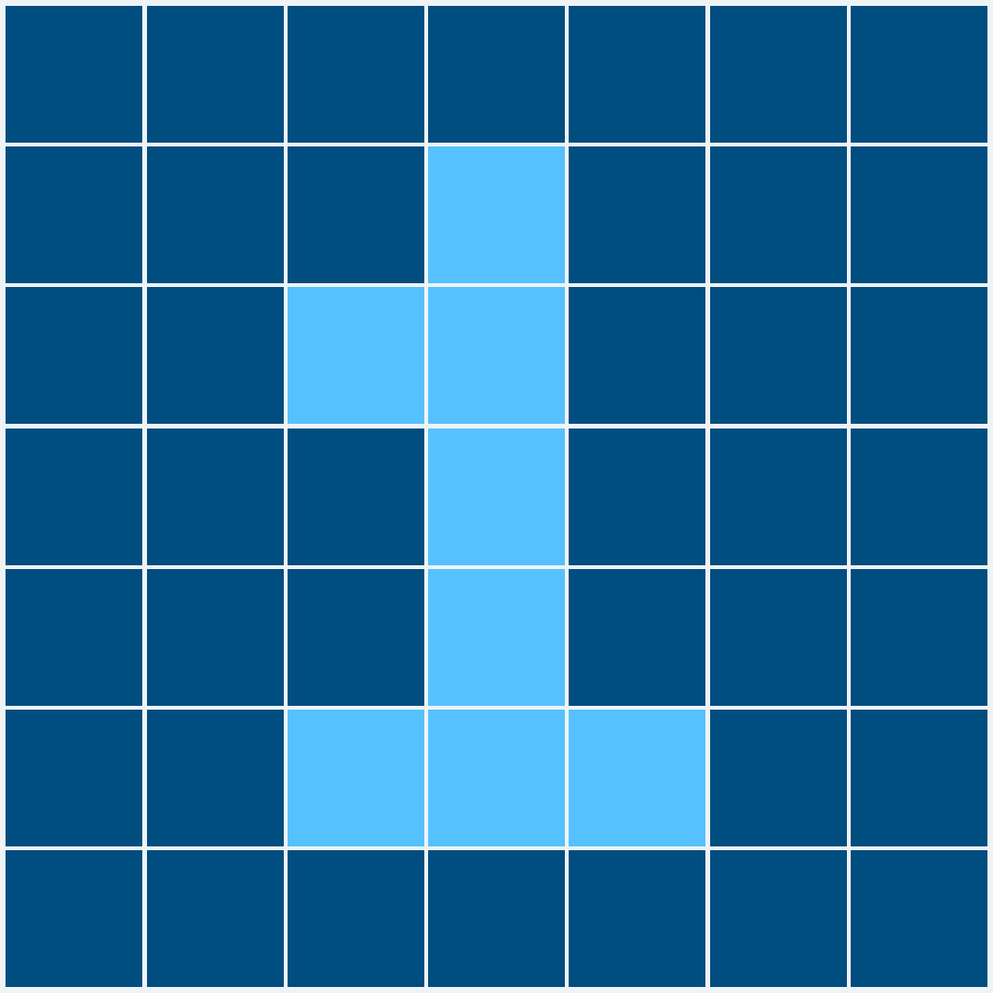




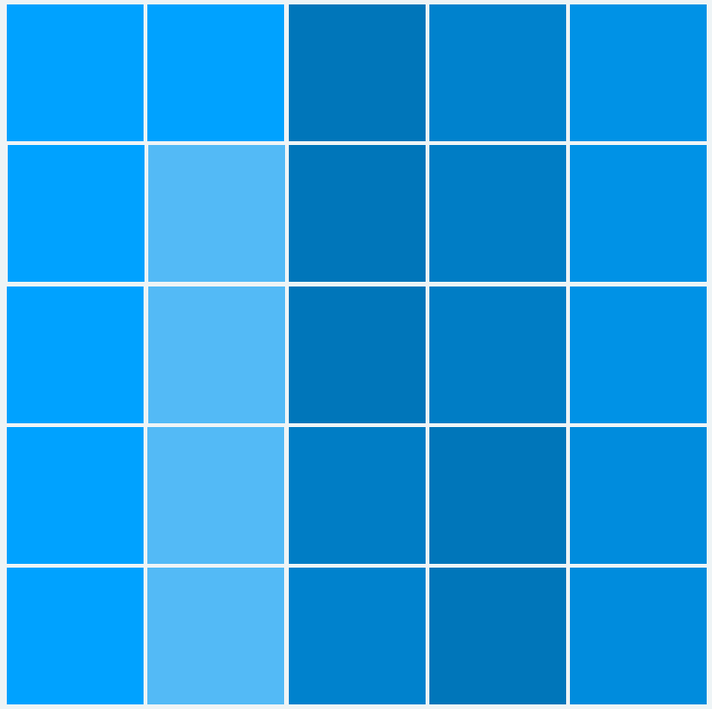
Pooling



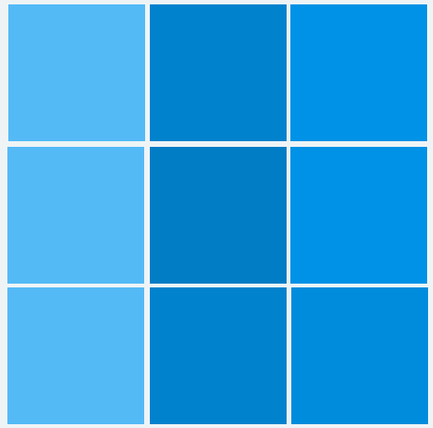


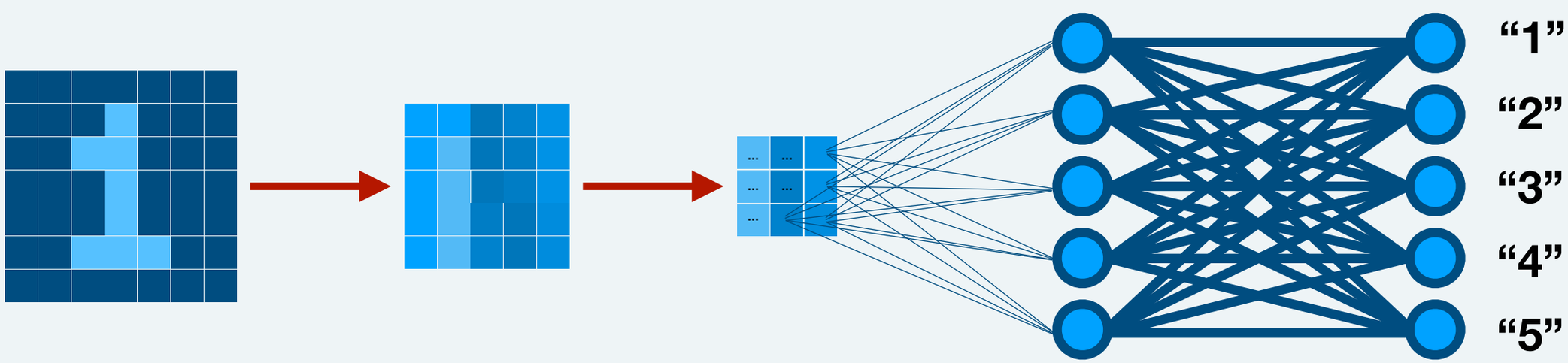


Convolution

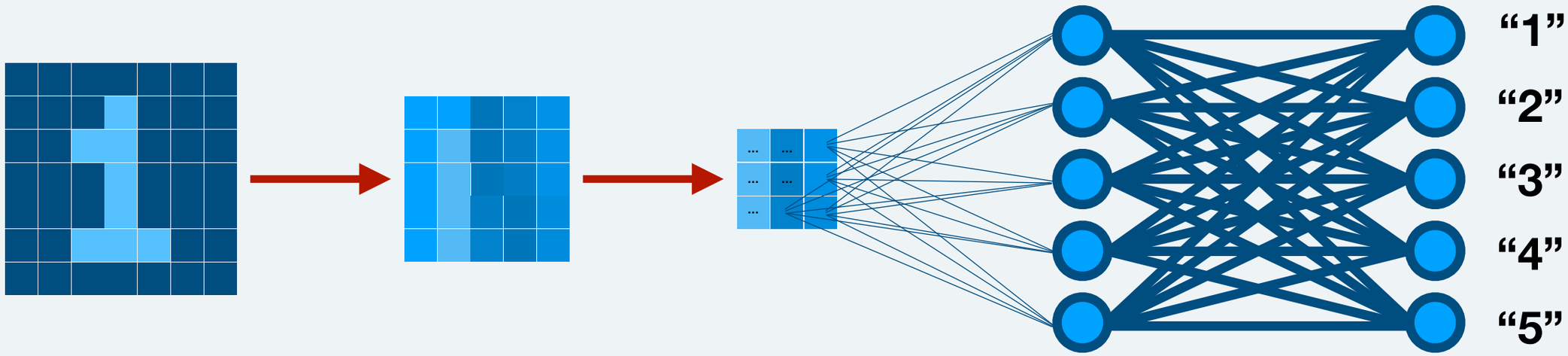


Pooling



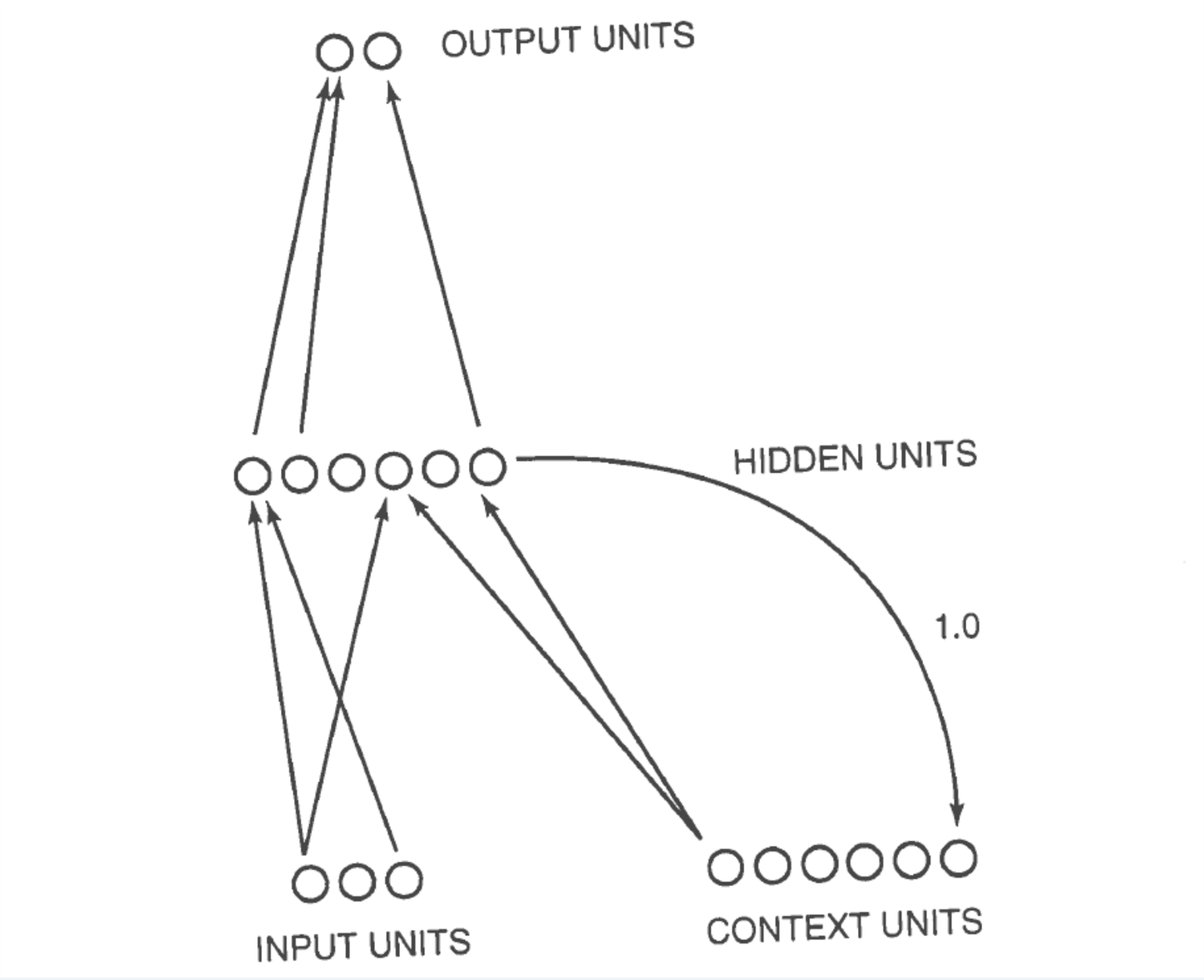

“1”
“2”
“3”
“4”
“5”

https://adamharley.com/nn_vis/



Connectionist networks

Recurrent network

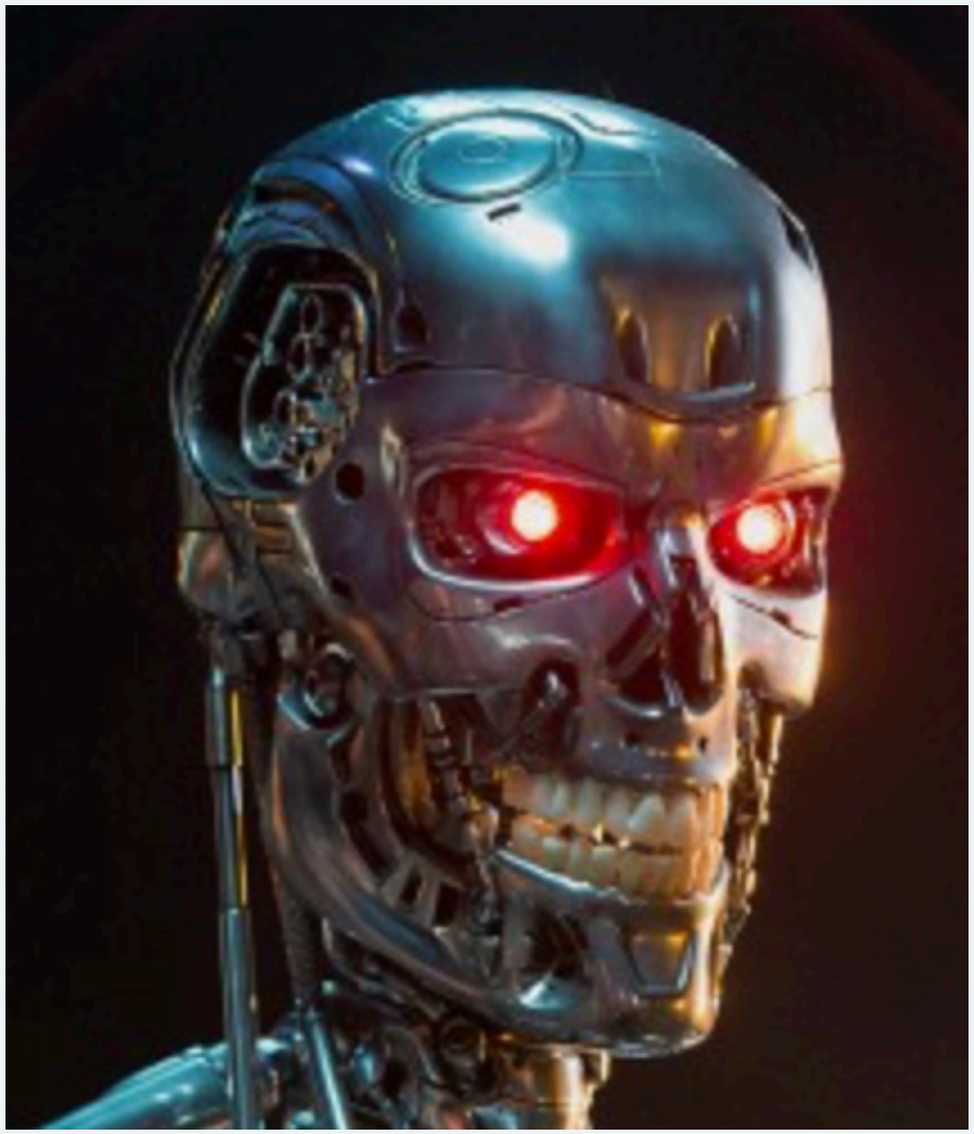


Revision

Let’s start by taking stock of what we’ve learnt about connectionism so far…

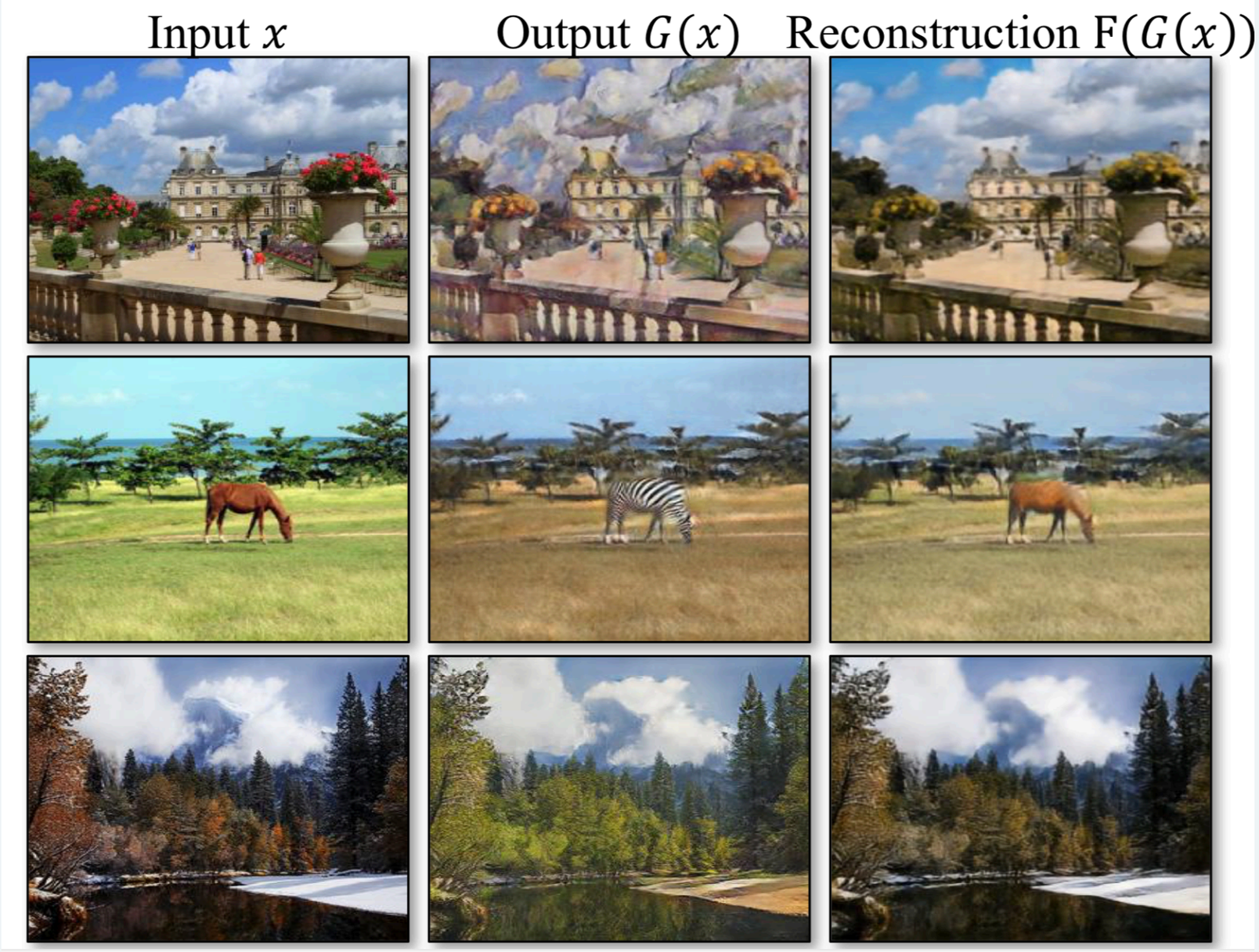
- the perceptron
- the perceptron learning rule
- linear separability
- limitations of single-layer perceptrons
- multi-layer neural networks
- gradient descent
- distributed representations
- fully connected, convolutional, recurrent network

Bad robots!



CycleGAN

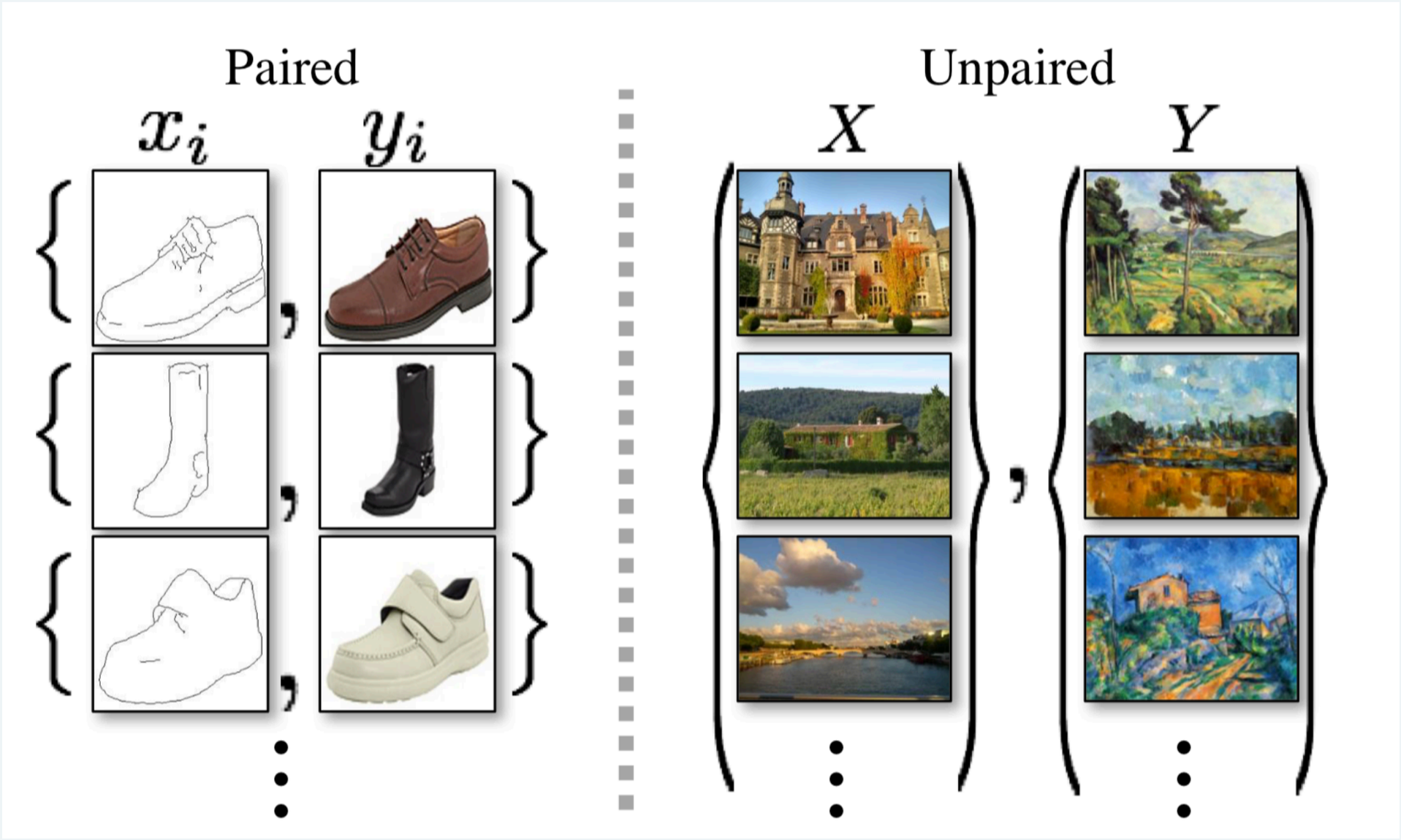
CycleGAN: a deep neural network for converting images from one domain into another



link

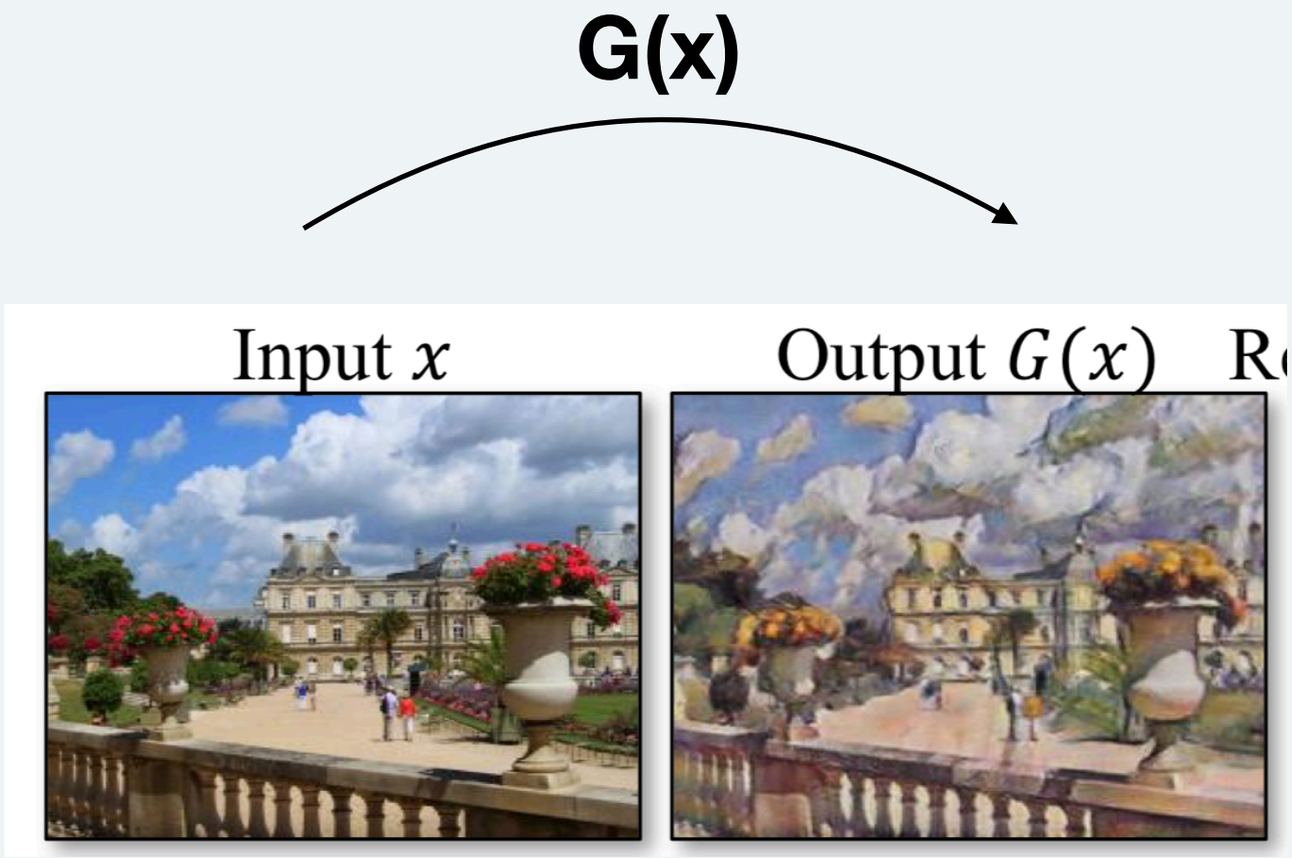
CycleGAN

CycleGAN does this without relying on paired images



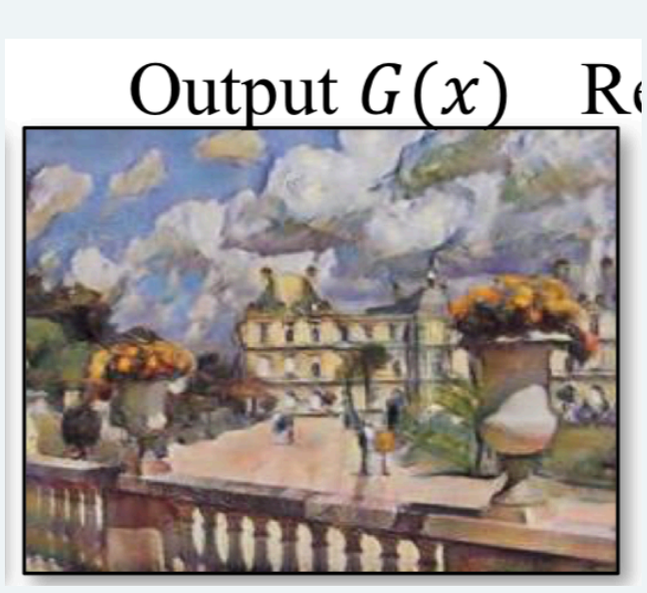
CycleGAN

first network: learns conversion G(x) from domain A to domain B



CycleGAN

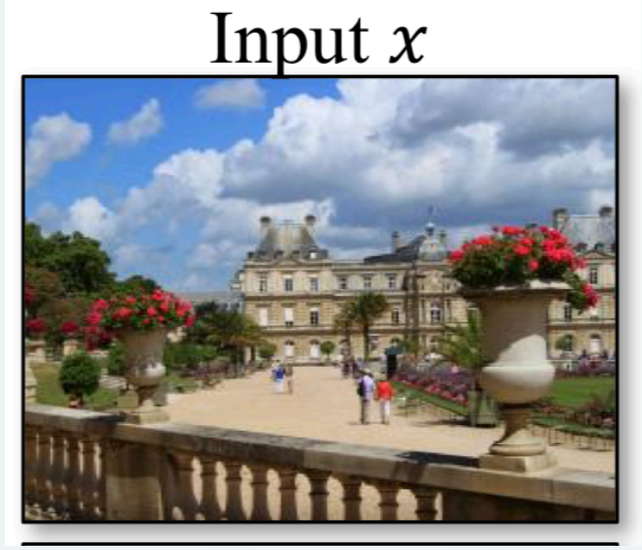
second network: tries to tell if G(x) is a ‘fake’ or a real image from Y

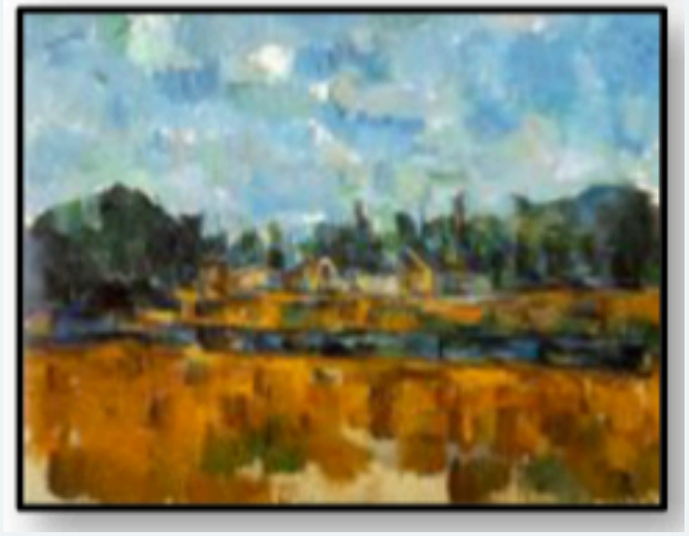


vs.

CycleGAN

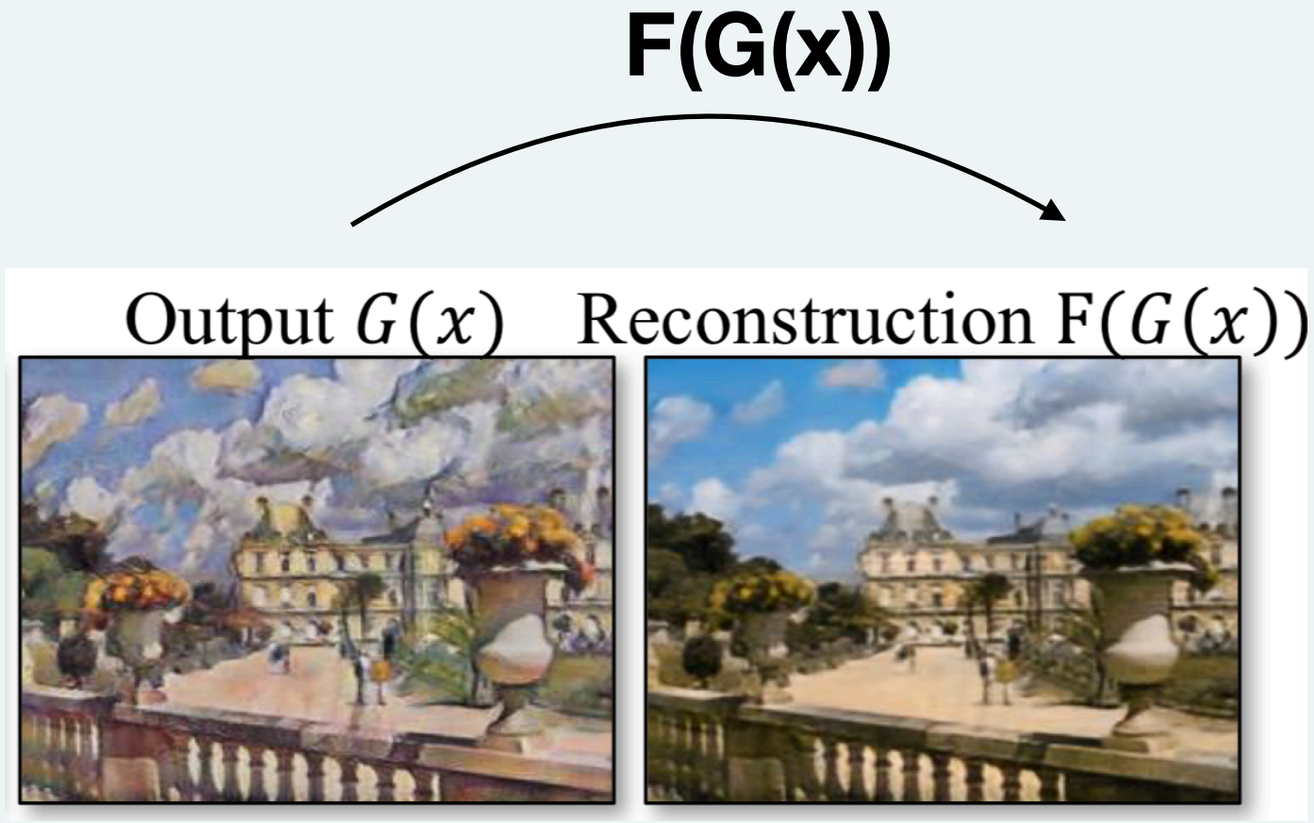
but, at this point, x could in principle be translated into *any* image from Y:





CycleGAN

third network: learns conversion F(y) from domain Y to domain X



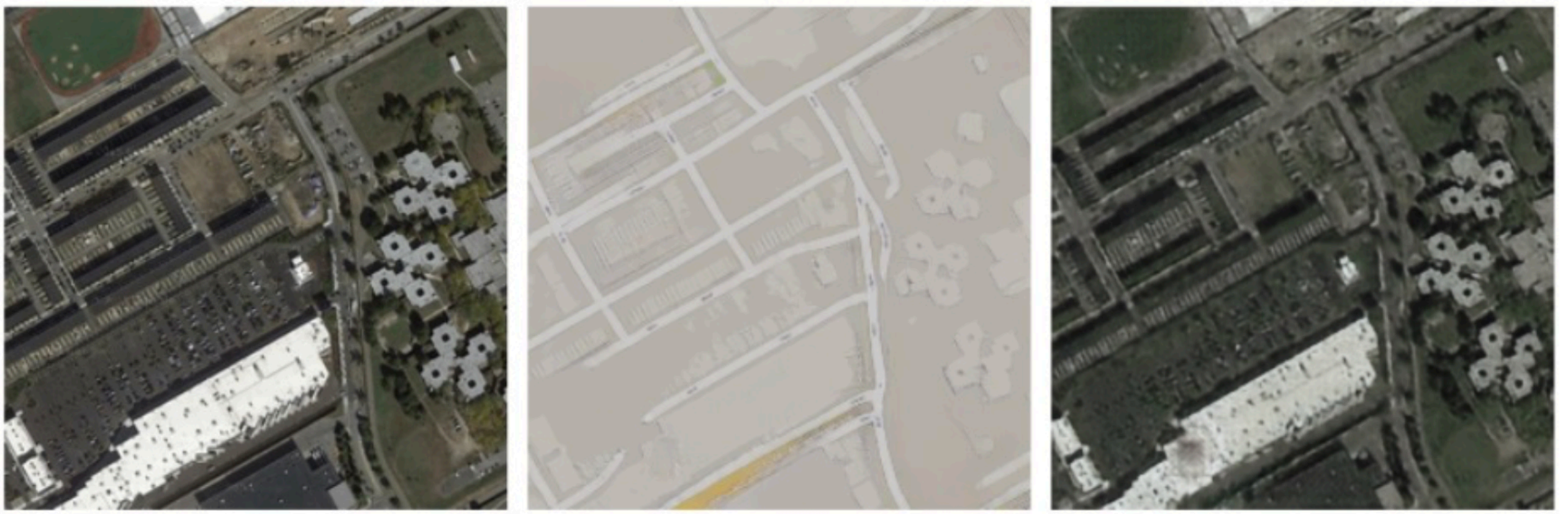
CycleGAN

the whole AI is constrained to minimise the difference between x and F(G(x))

diff(x, F(G(x)))

A connectionist network for creating maps

CycleGAN

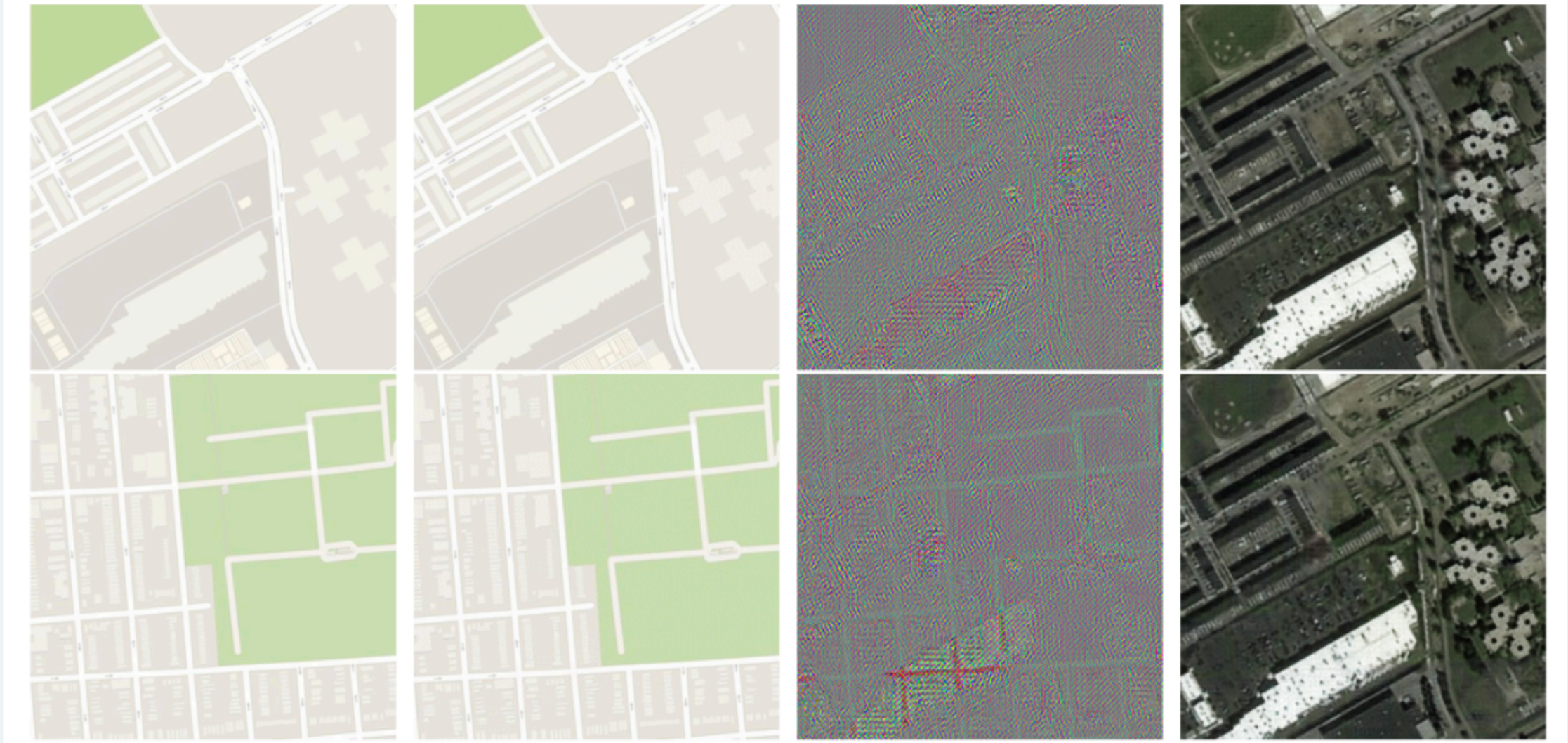


(a) Aerial photograph: x. (b) Generated map: Fx. (c) Aerial reconstruction: GFx.

link

A connectionist network for creating maps

CycleGAN



(a) Source map: y_0. (b) Crafted map: y^*. (c) Difference: $y^* - y_0$. (d) Reconstruction: Gy^*.

link

“But in fact this occurrence, far from illustrating some kind of malign intelligence inherent to AI, simply reveals a problem with computers that has existed since they were invented: they do exactly what you tell them to do.”

source: TechCrunch, link