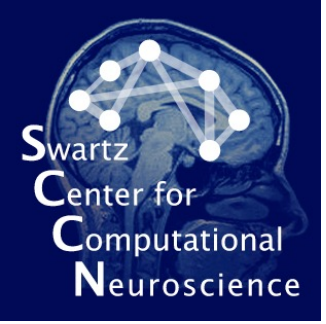


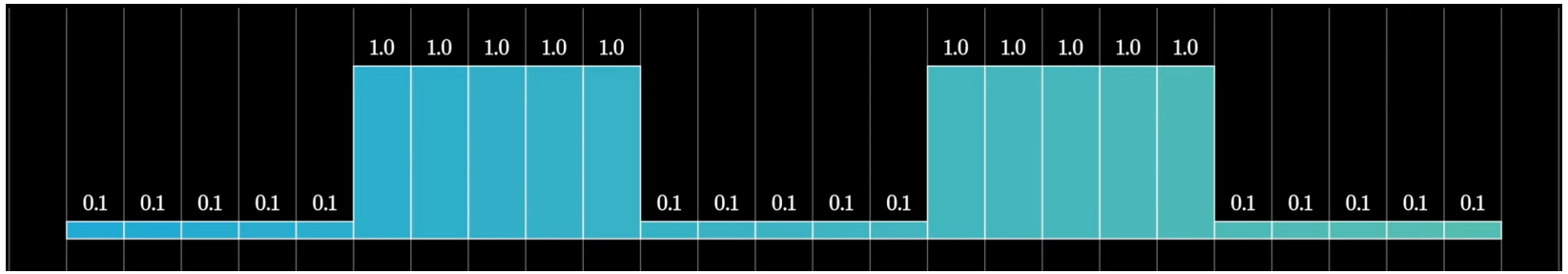
UC San Diego

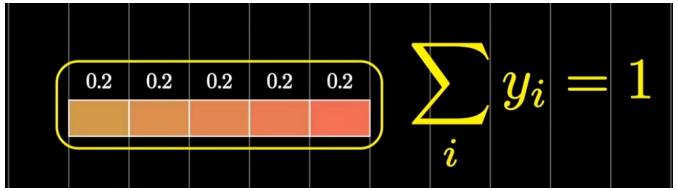
Université Paul Sabatier
TOULOUSE III

Deep Learning applied to EEG

Dung “Young” Truong & Arnaud Delorme

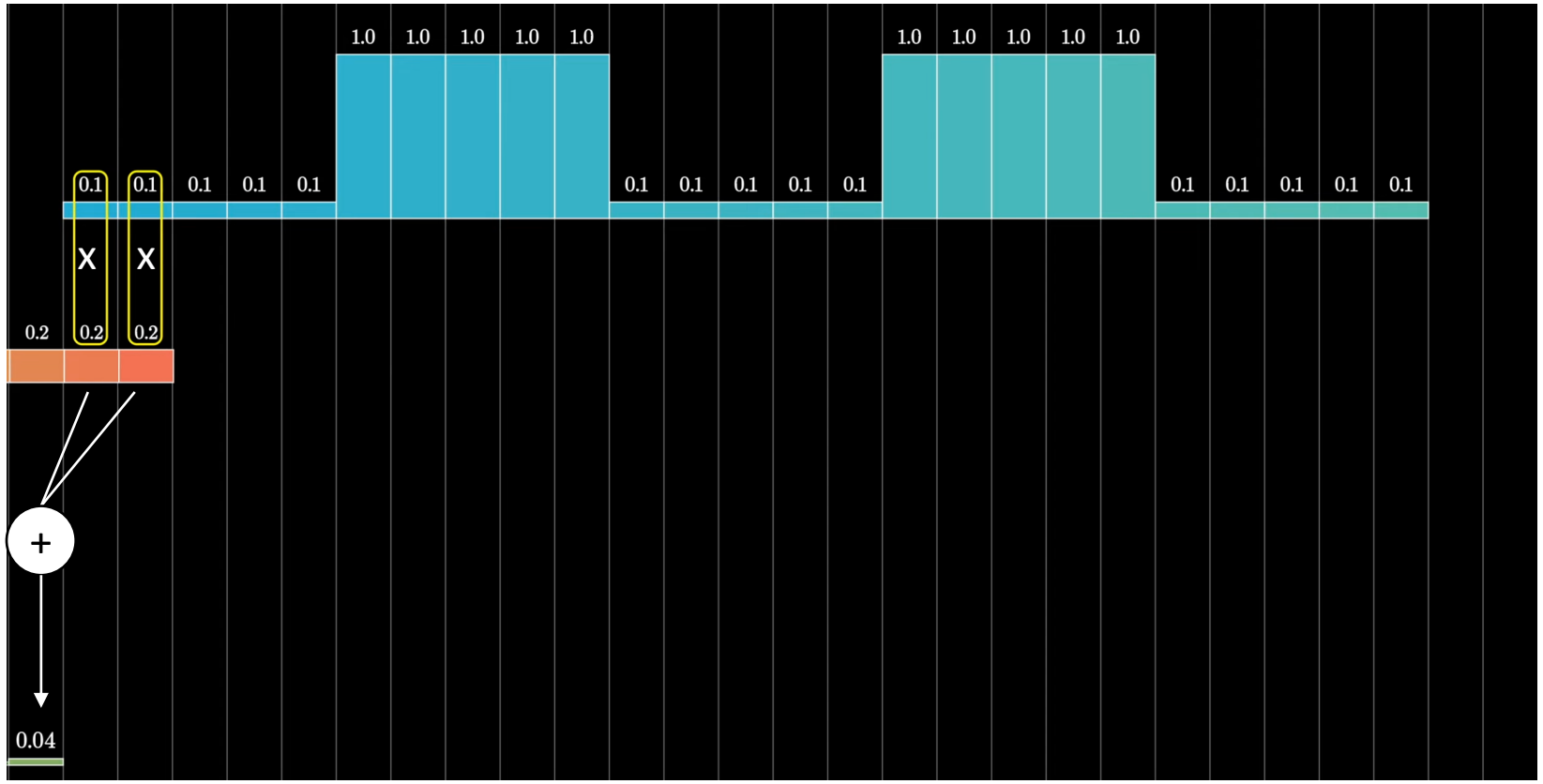
What kind of operation?





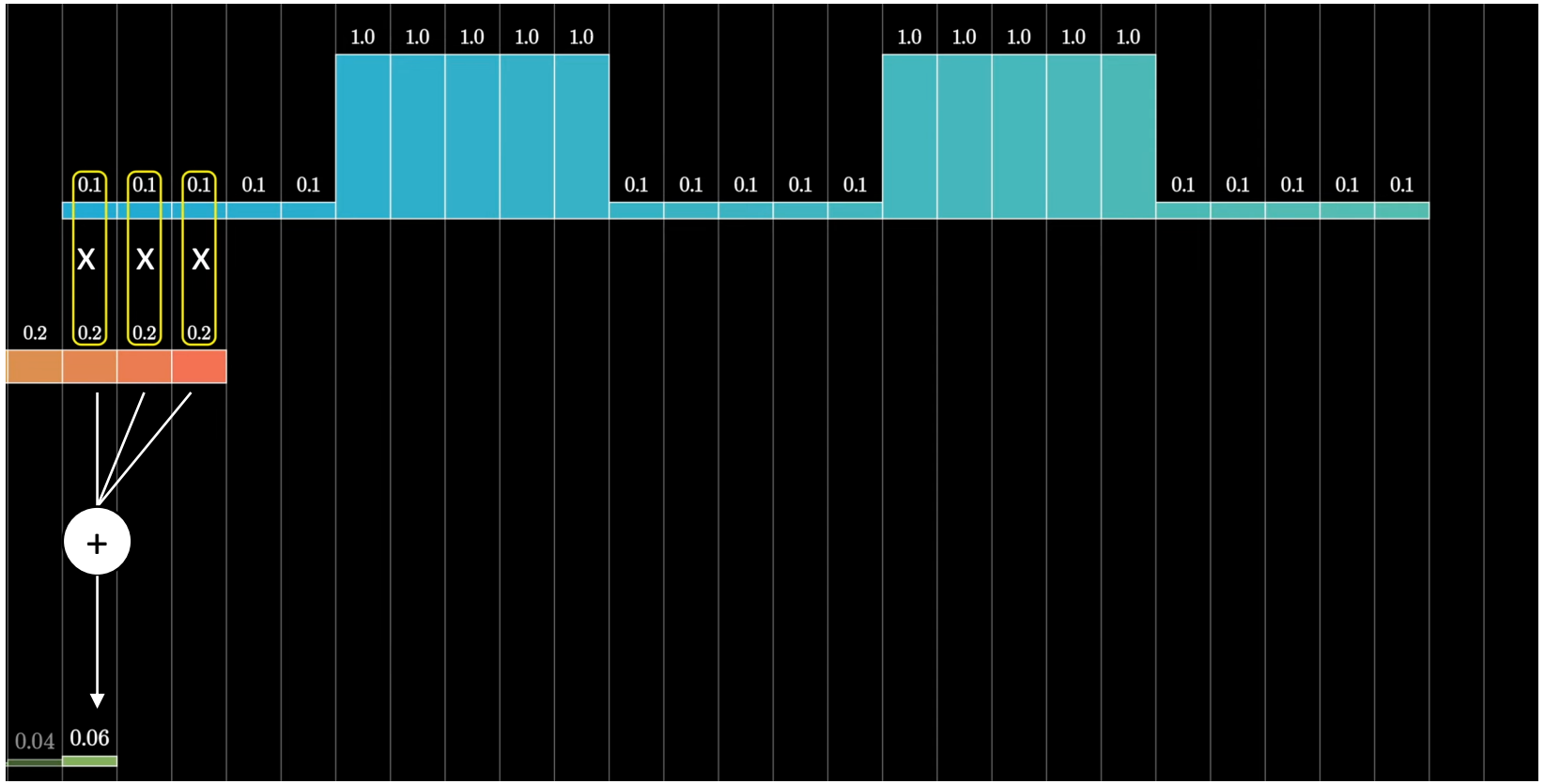
https://youtu.be/KuXjwB4LzSA

What kind of operation?



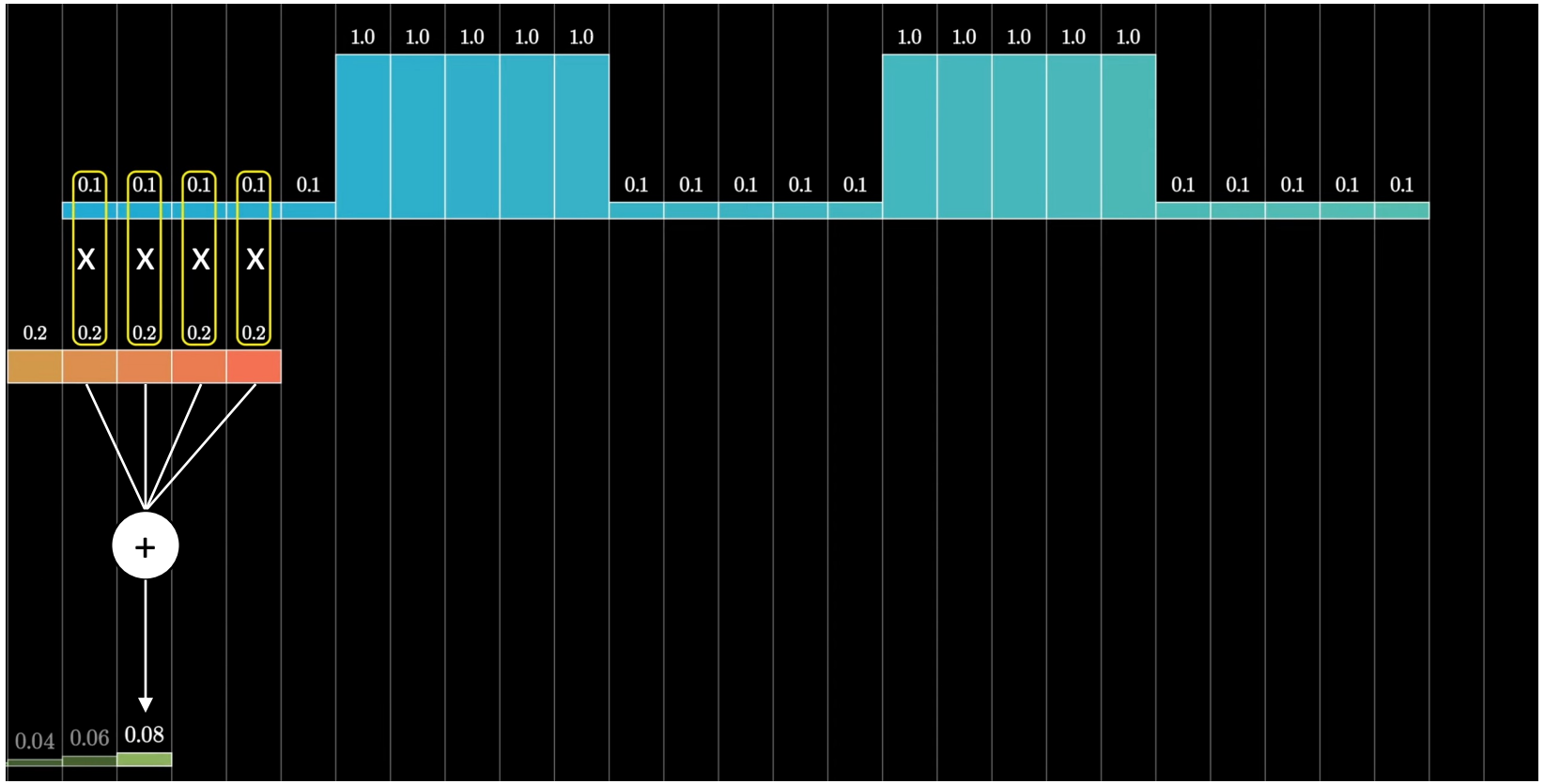
https://youtu.be/KuXjwB4LzSA

What kind of operation?



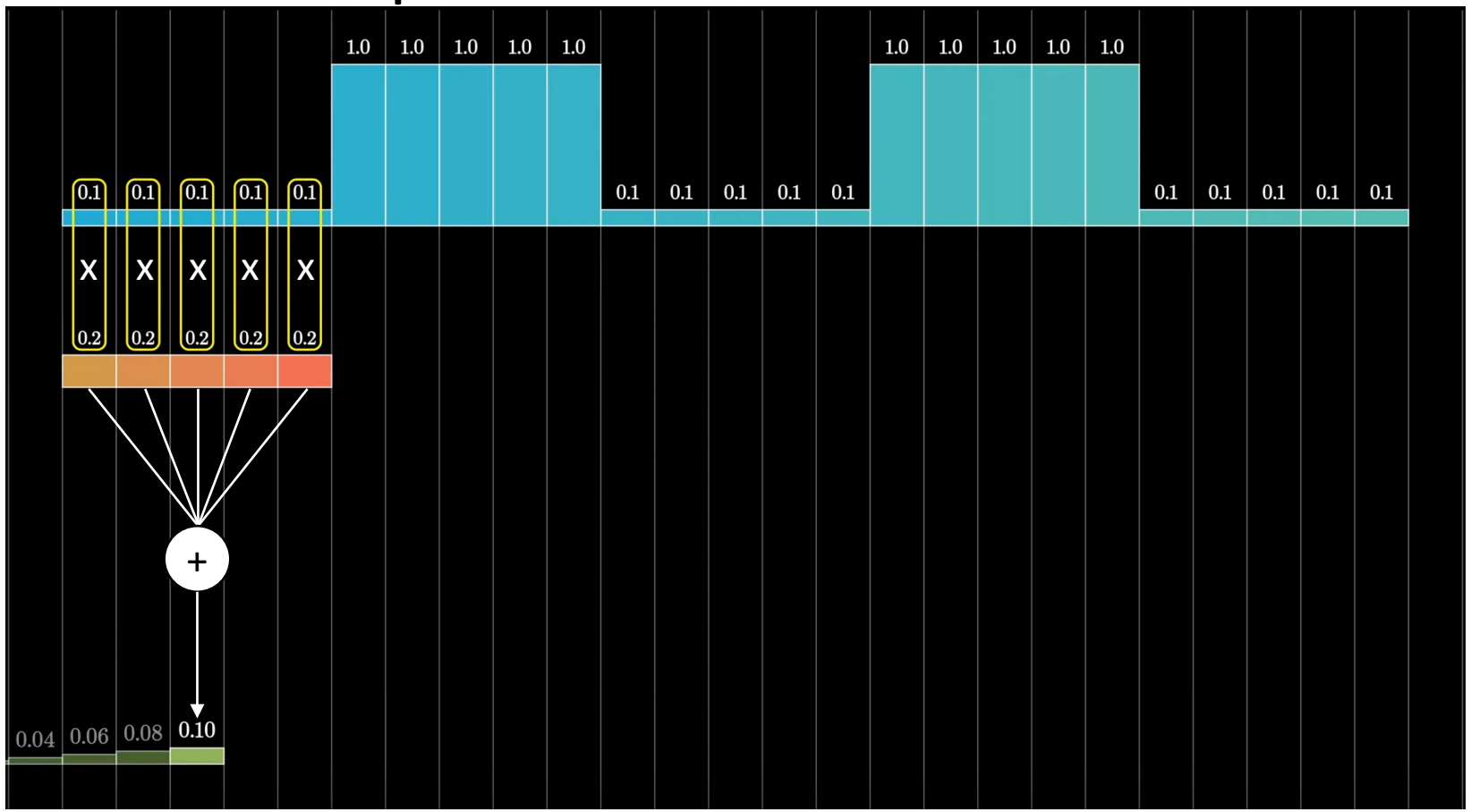
https://youtu.be/KuXjwB4LzSA

What kind of operation?



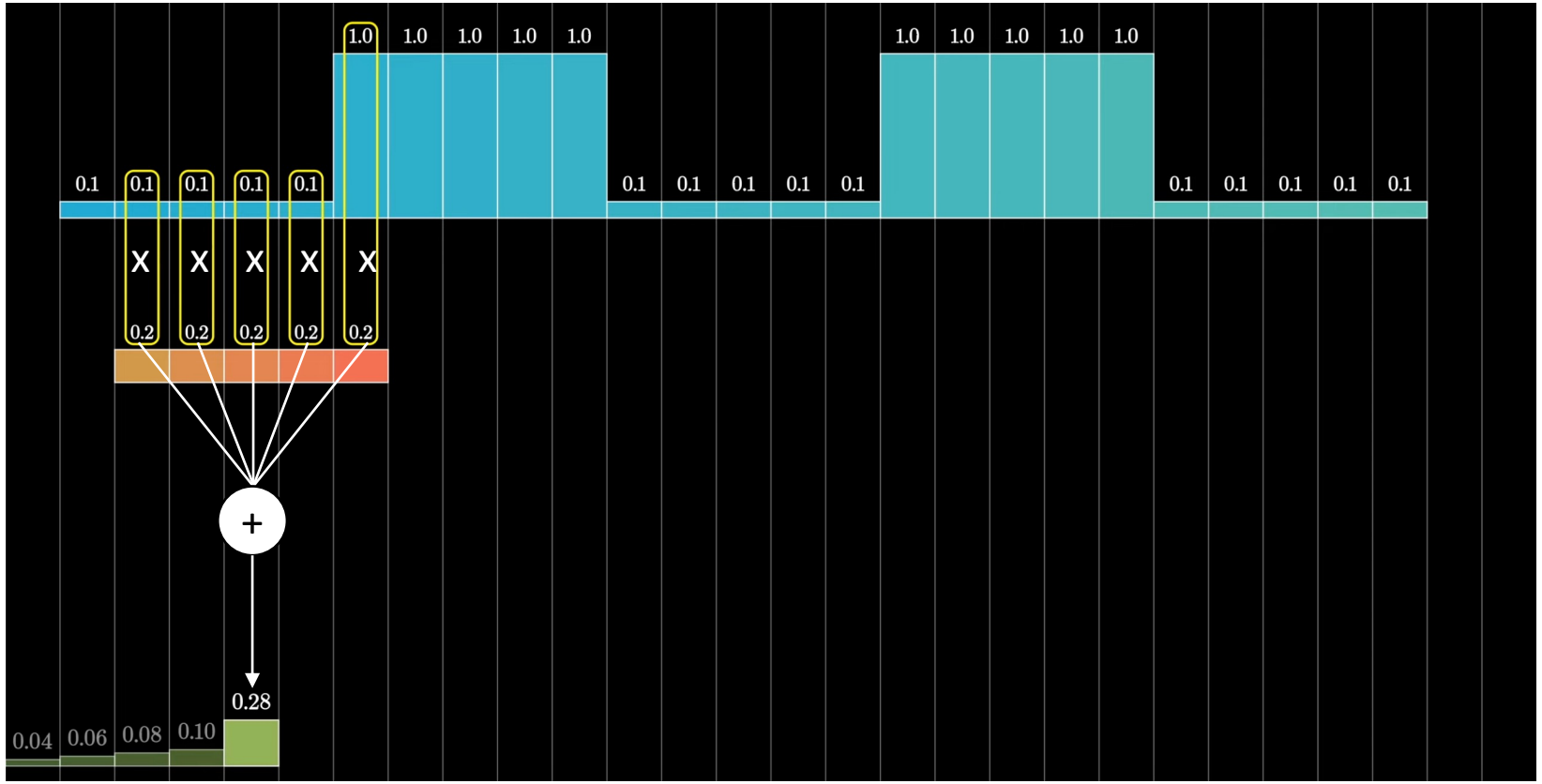
https://youtu.be/KuXjwB4LzSA

What kind of operation?



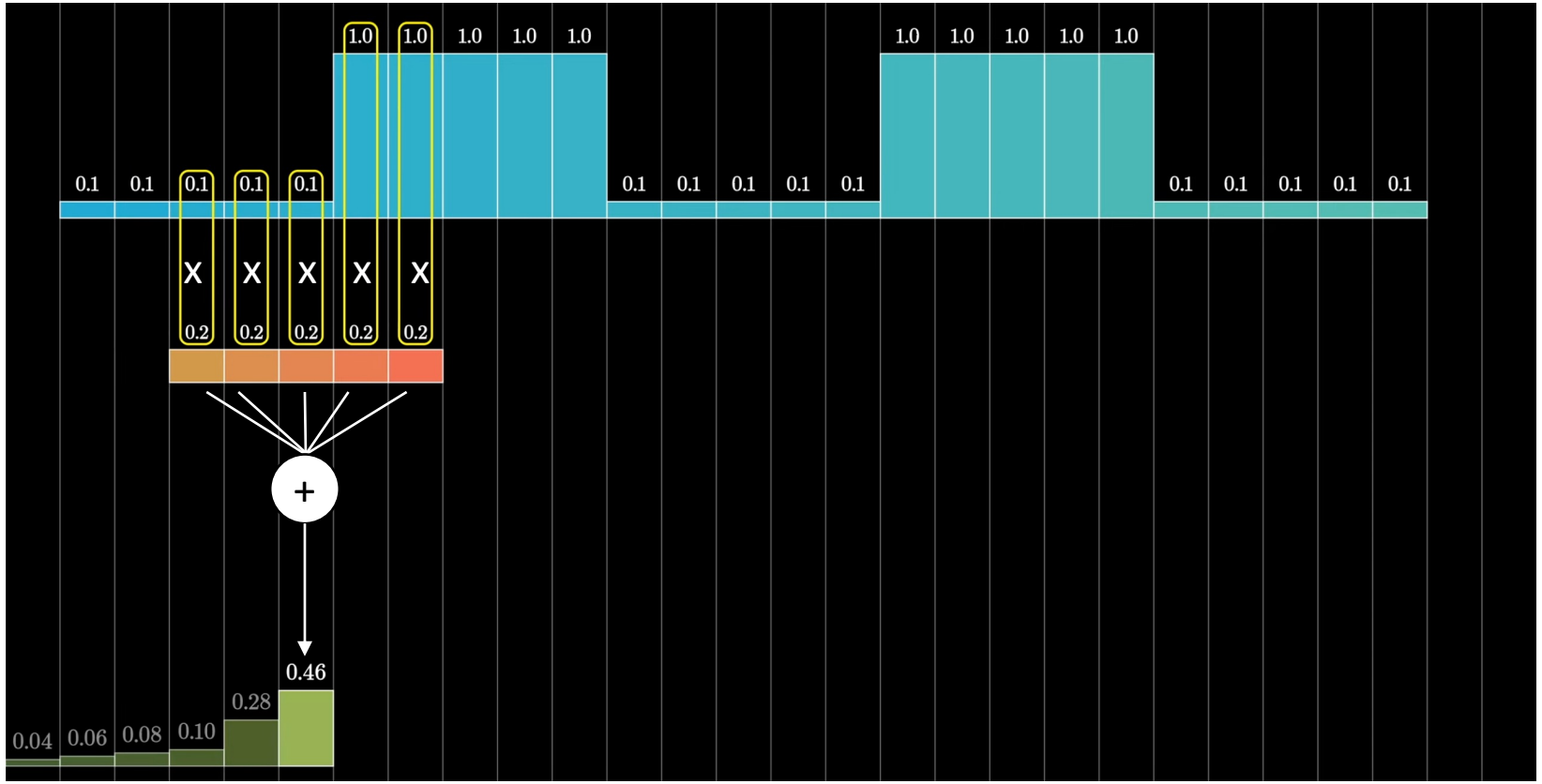
https://youtu.be/KuXjwB4LzSA

What kind of operation?



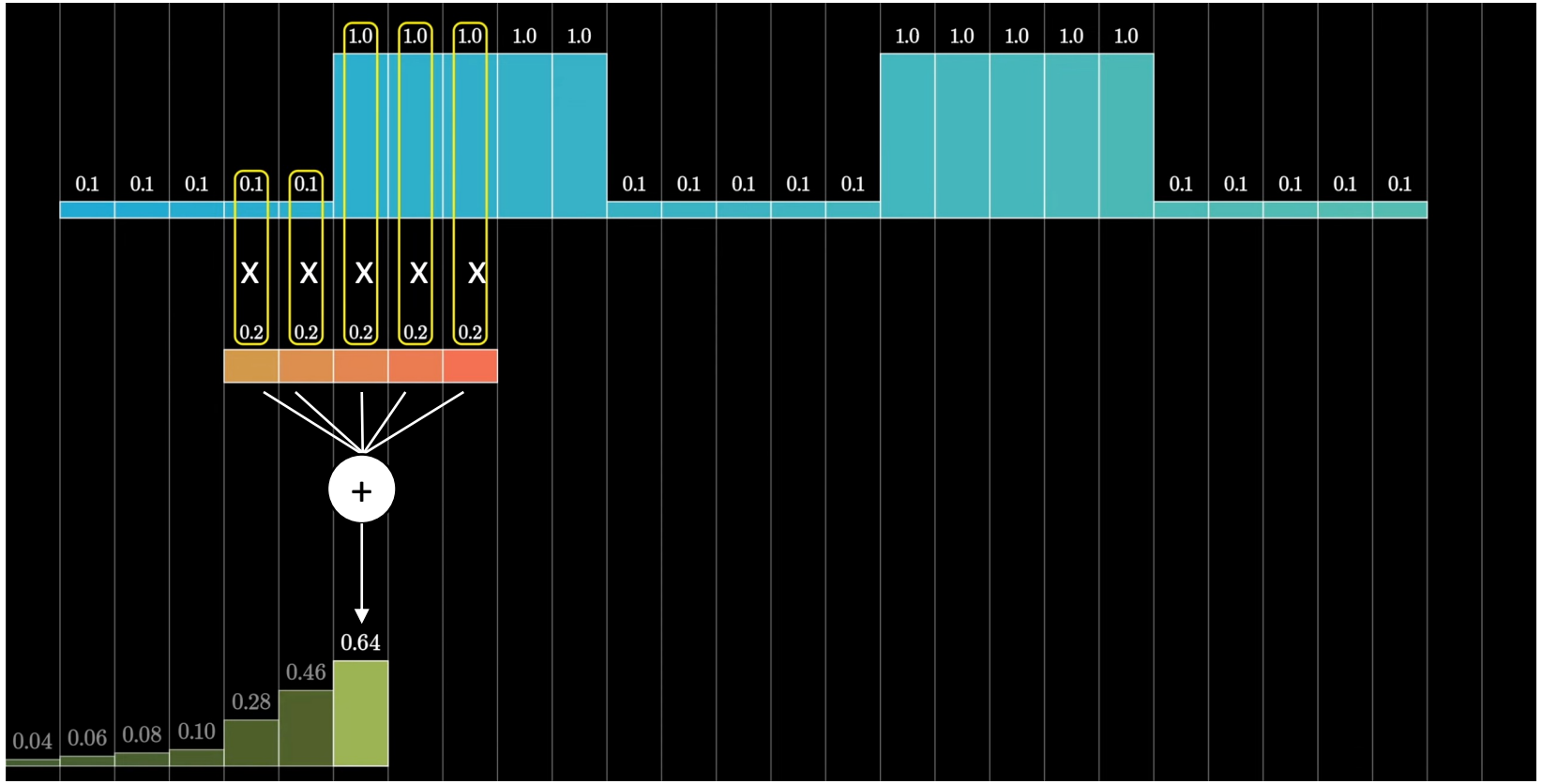
https://youtu.be/KuXjwB4LzSA

What kind of operation?



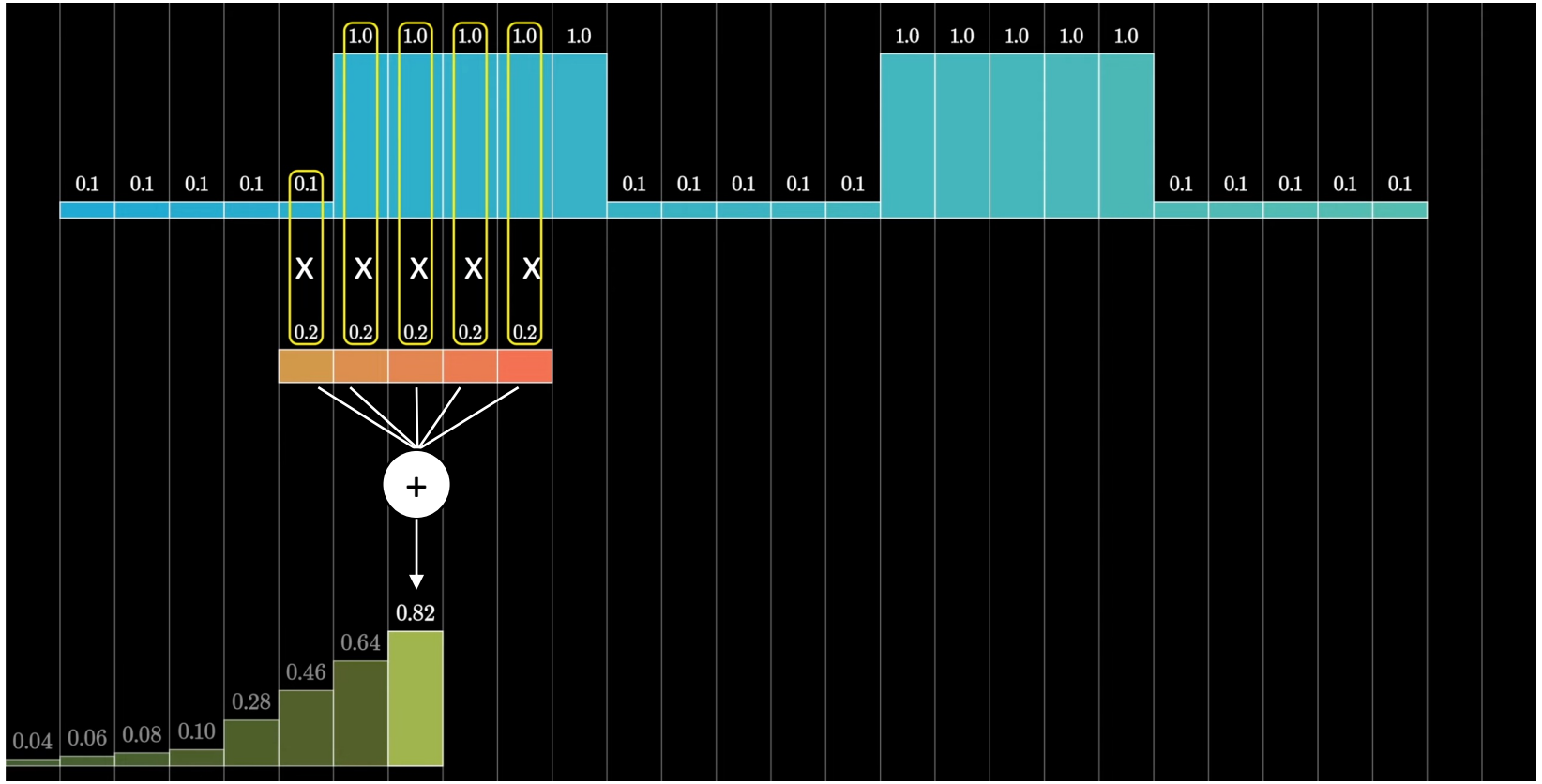
https://youtu.be/KuXjwB4LzSA

What kind of operation?



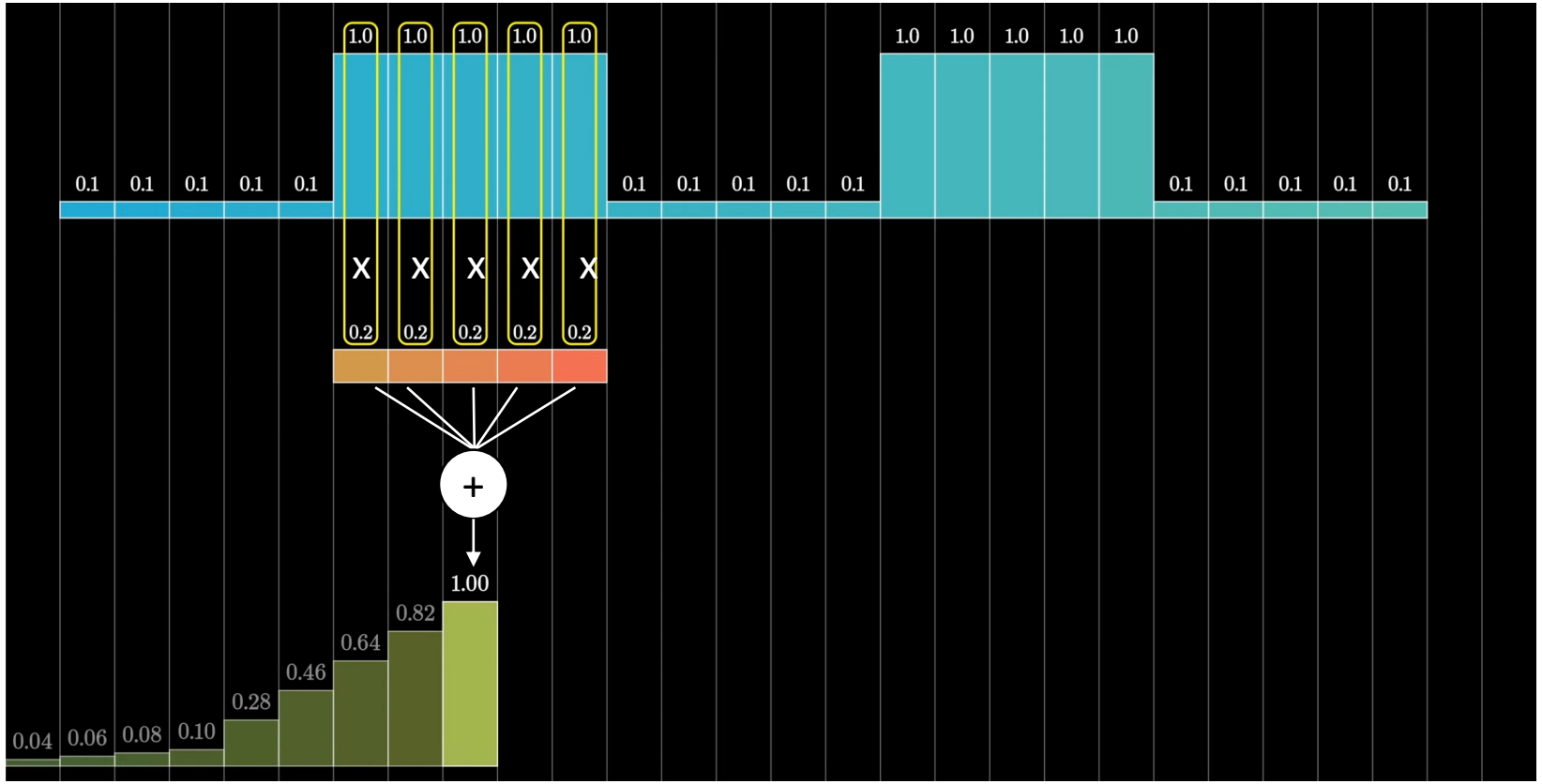
https://youtu.be/KuXjwB4LzSA

What kind of operation?



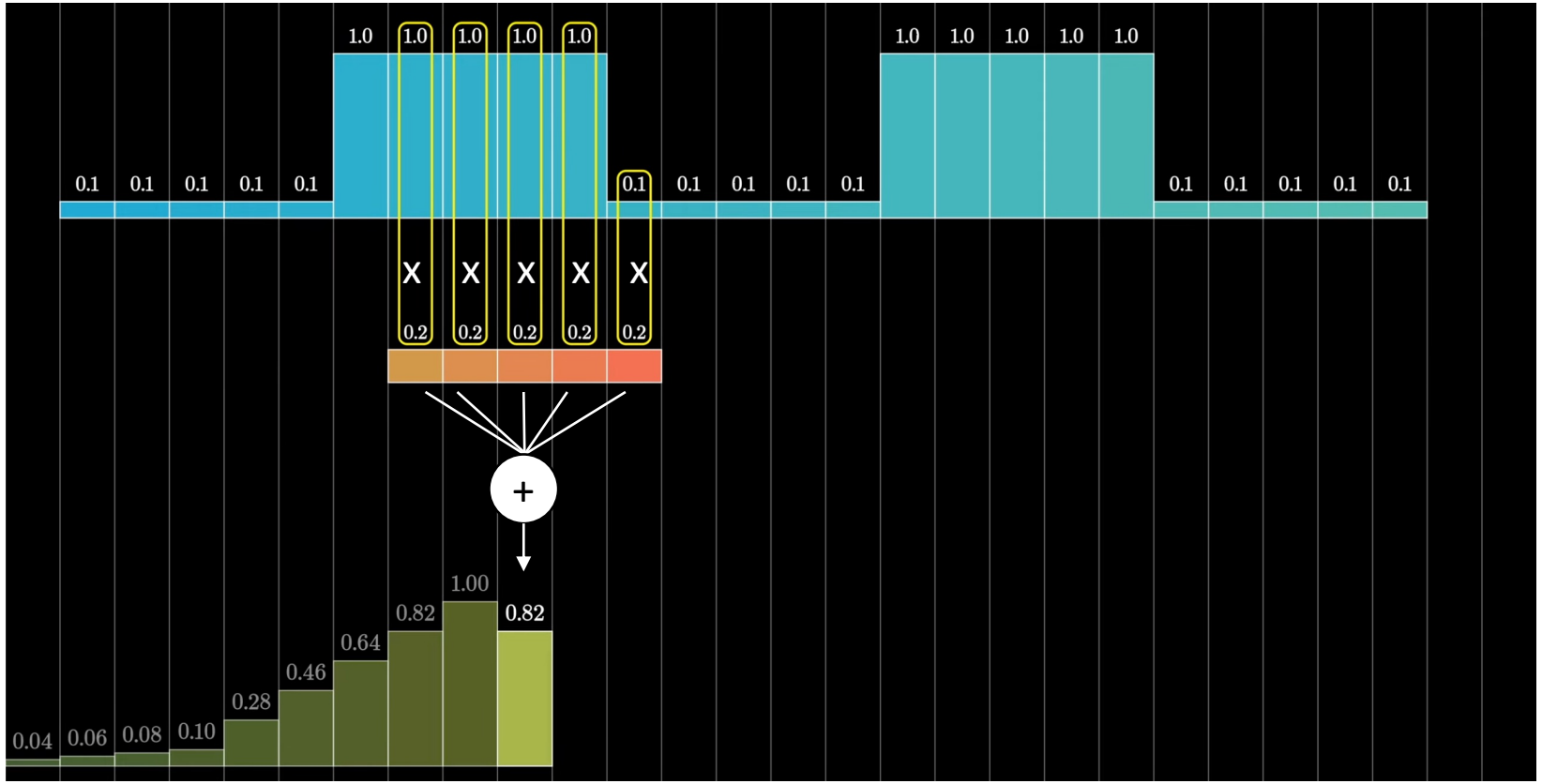
https://youtu.be/KuXjwB4LzSA

What kind of operation?



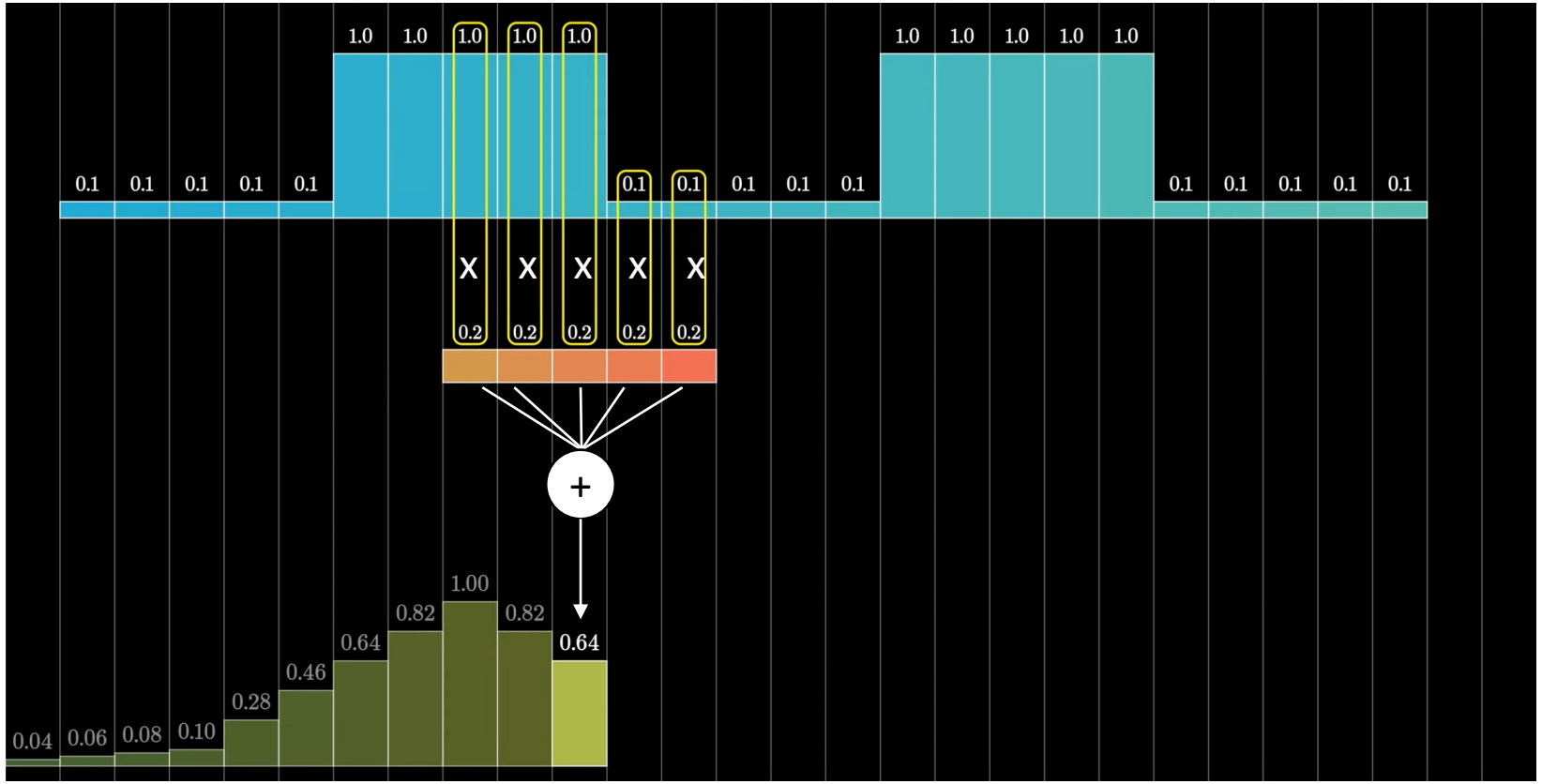
https://youtu.be/KuXjwB4LzSA

What kind of operation?



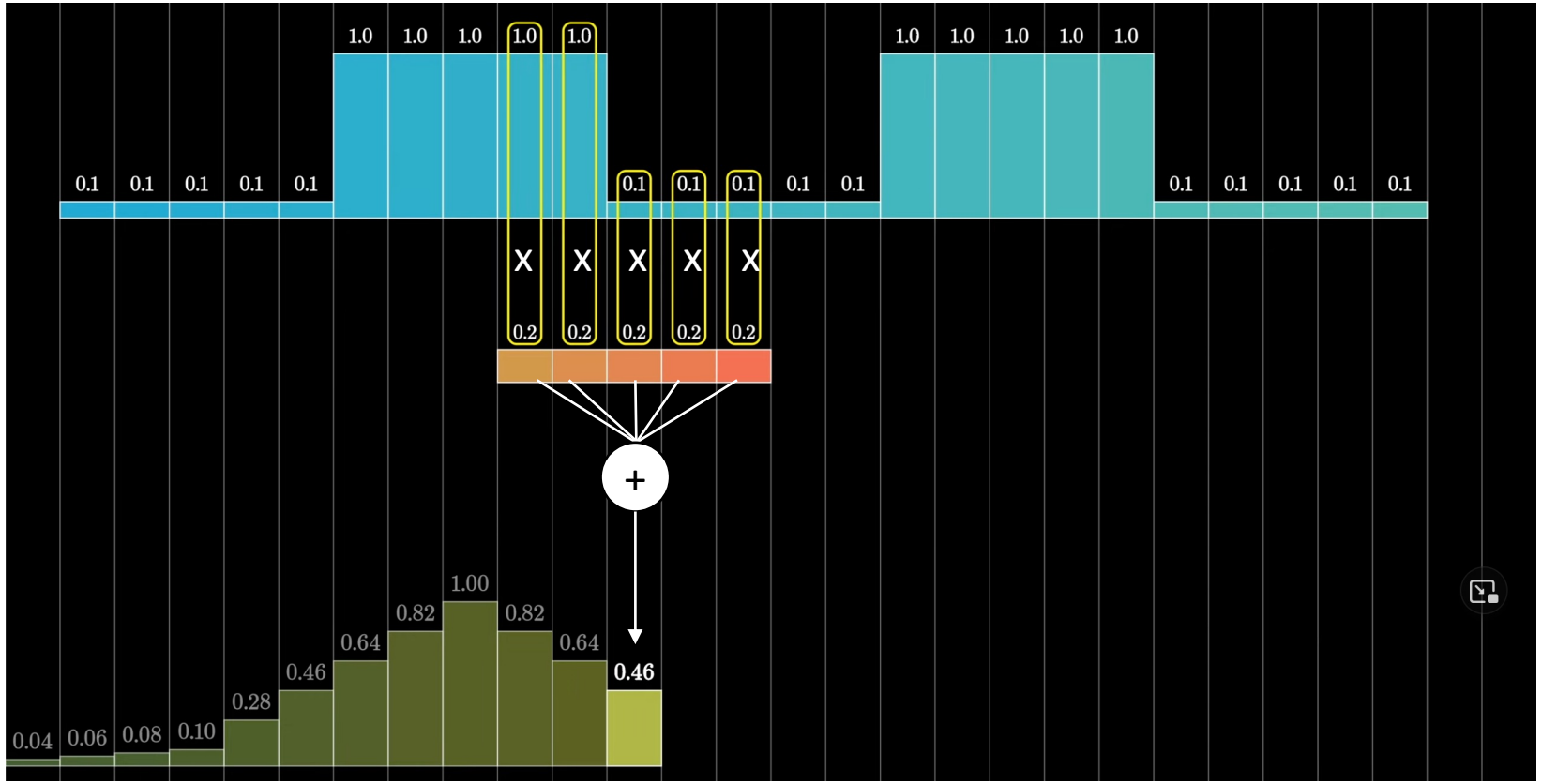
https://youtu.be/KuXjwB4LzSA

What kind of operation?



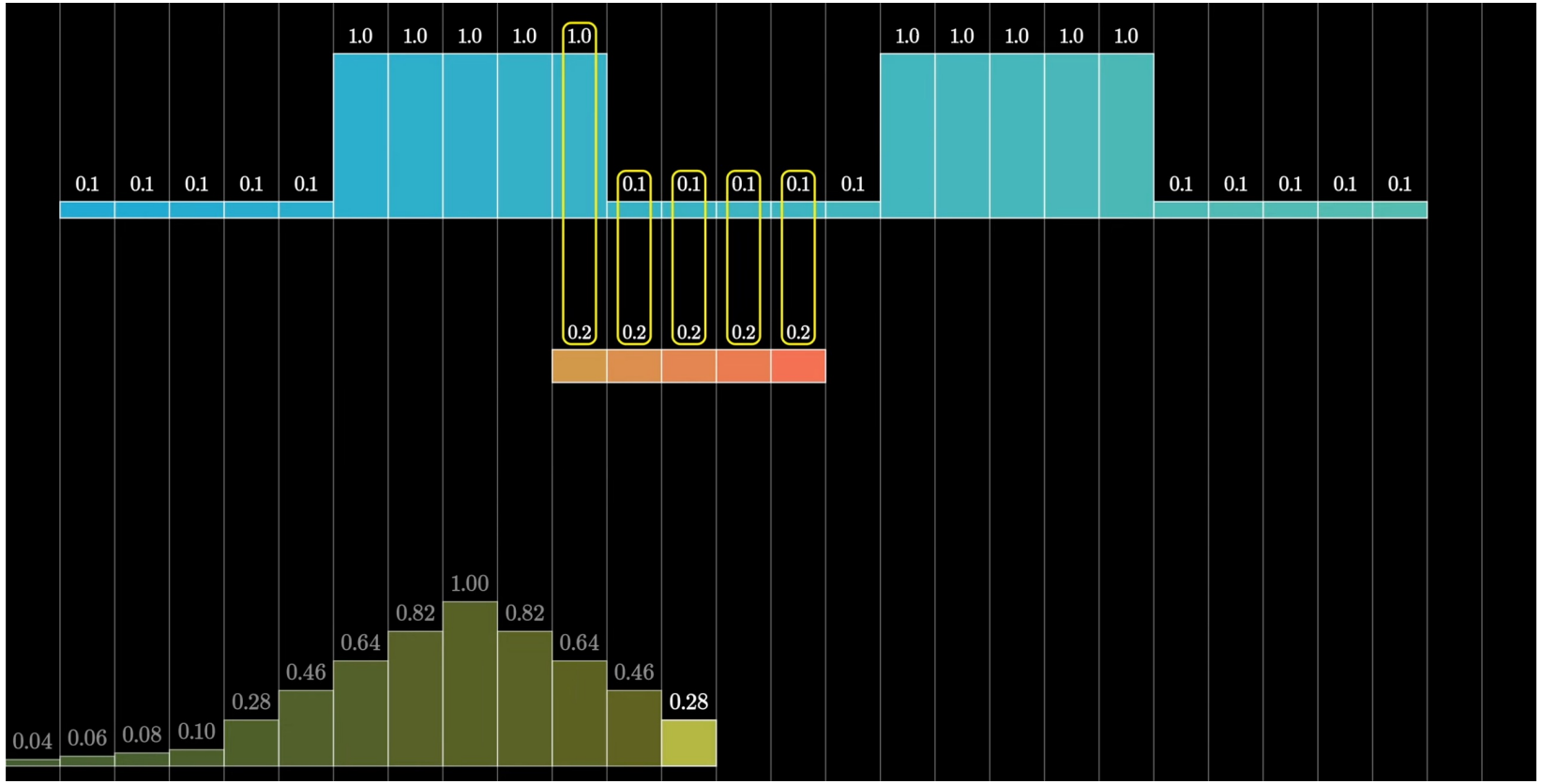
https://youtu.be/KuXjwB4LzSA

What kind of operation?



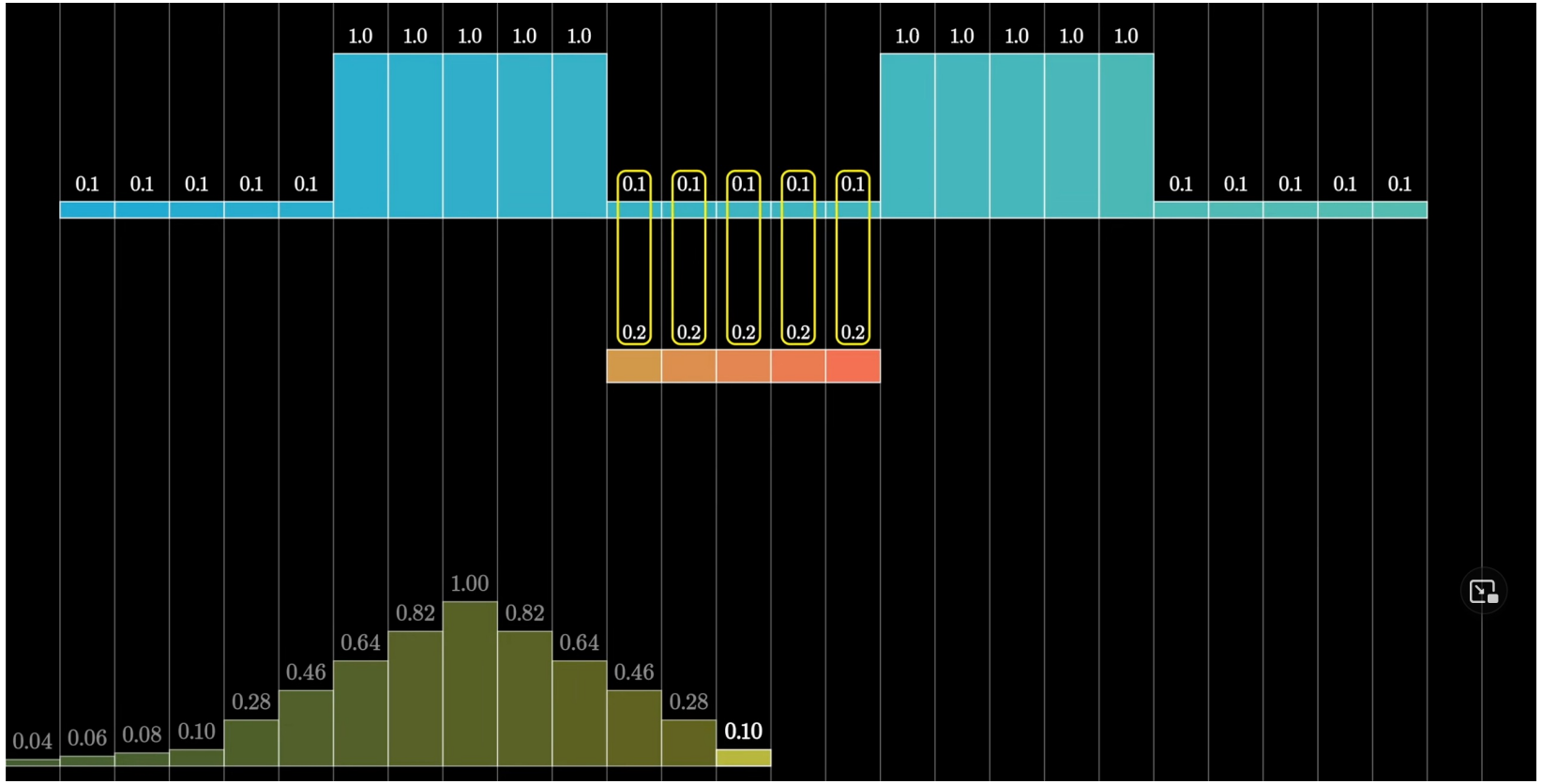
https://youtu.be/KuXjwB4LzSA

What kind of operation?



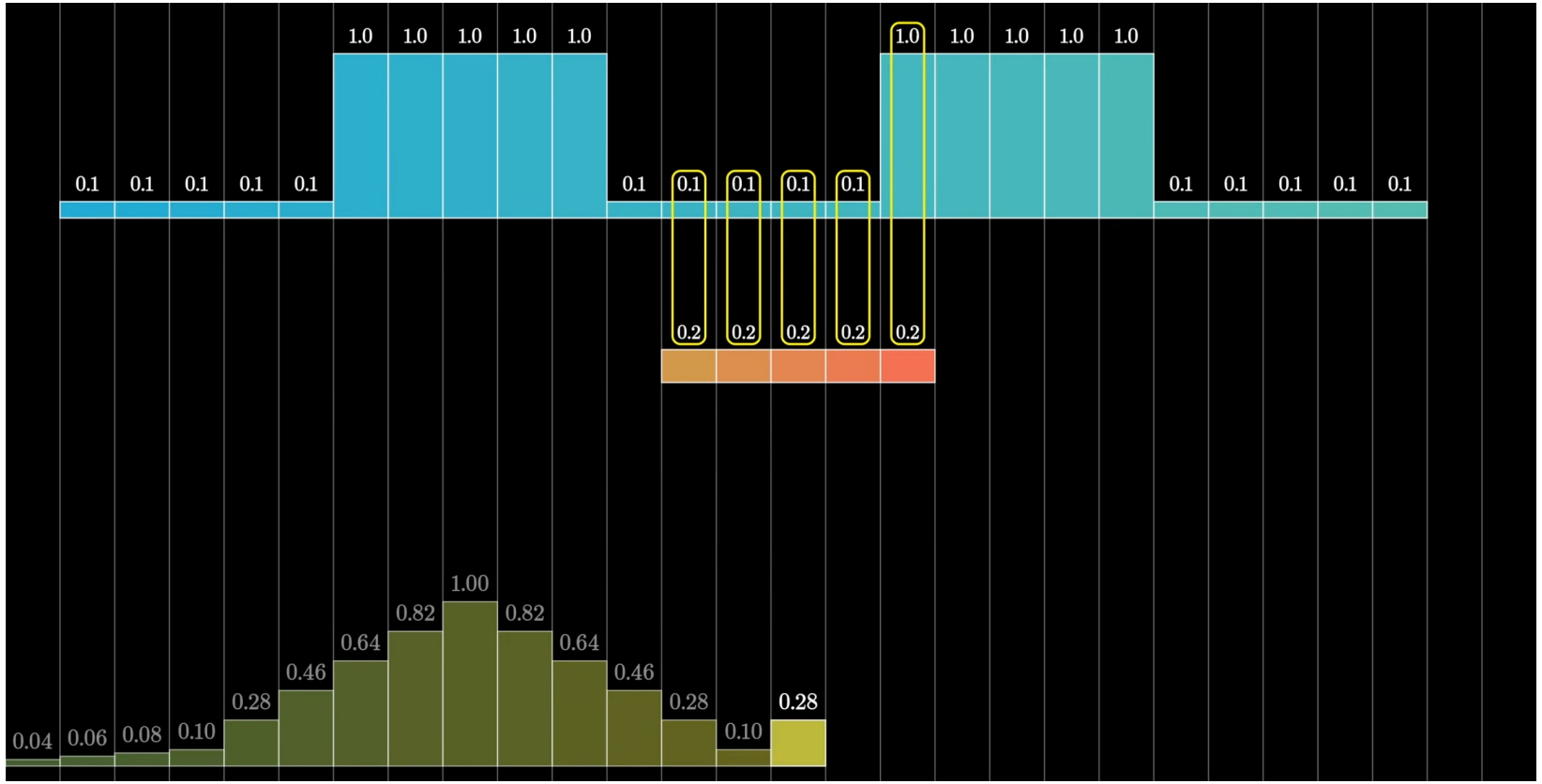
https://youtu.be/KuXjwB4LzSA

What kind of operation?



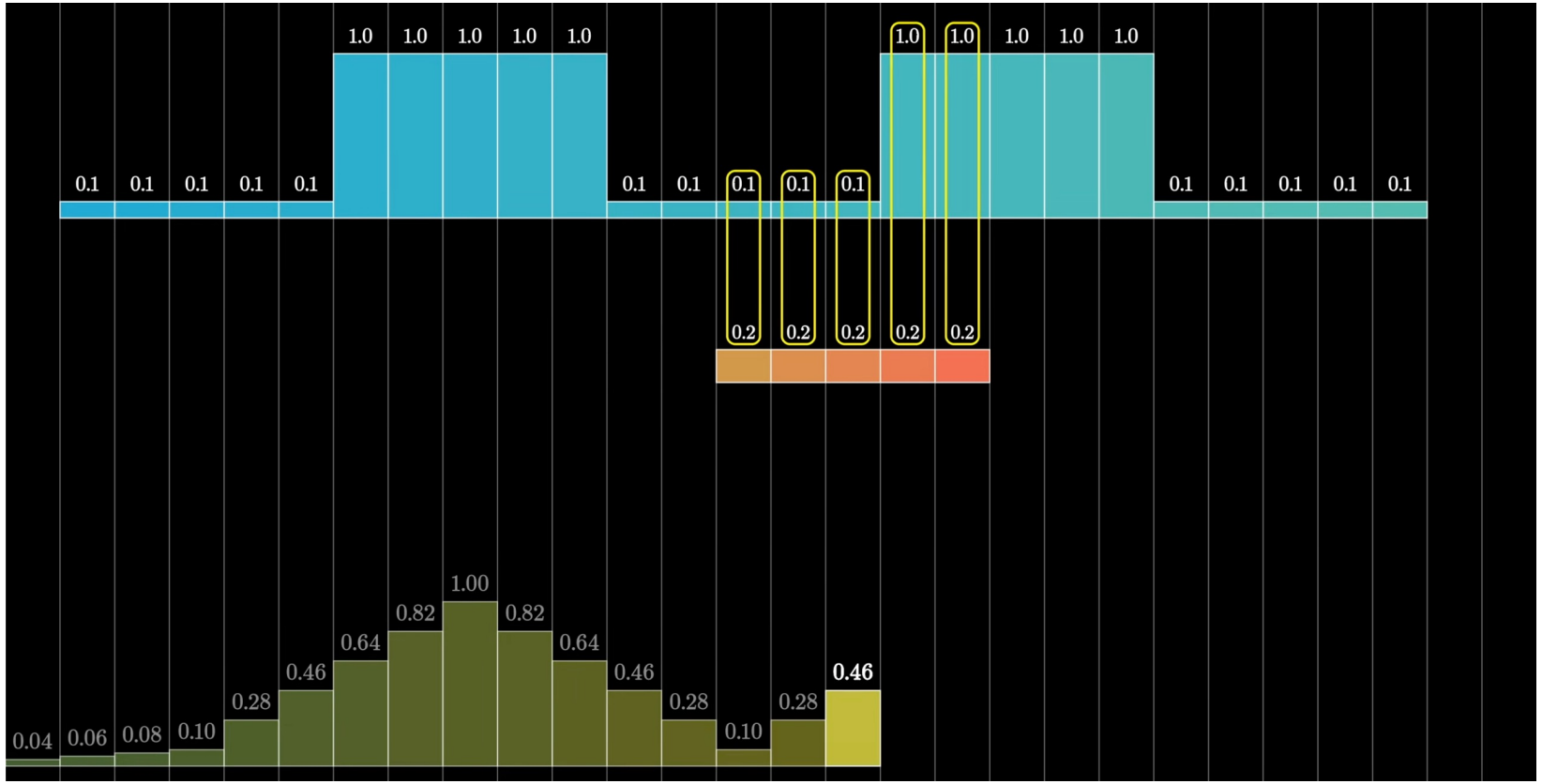
https://youtu.be/KuXjwB4LzSA

What kind of operation?



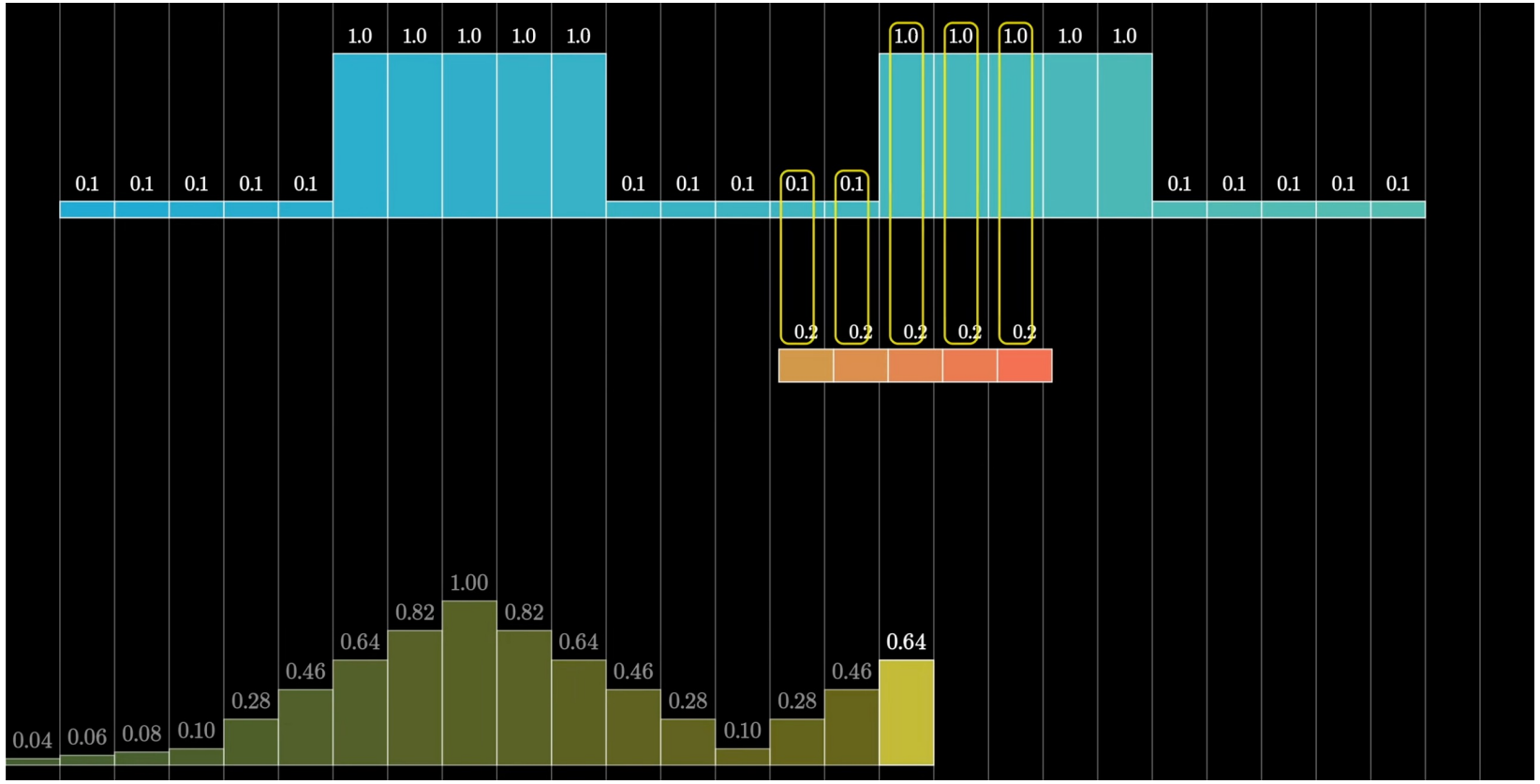
https://youtu.be/KuXjwB4LzSA

What kind of operation?



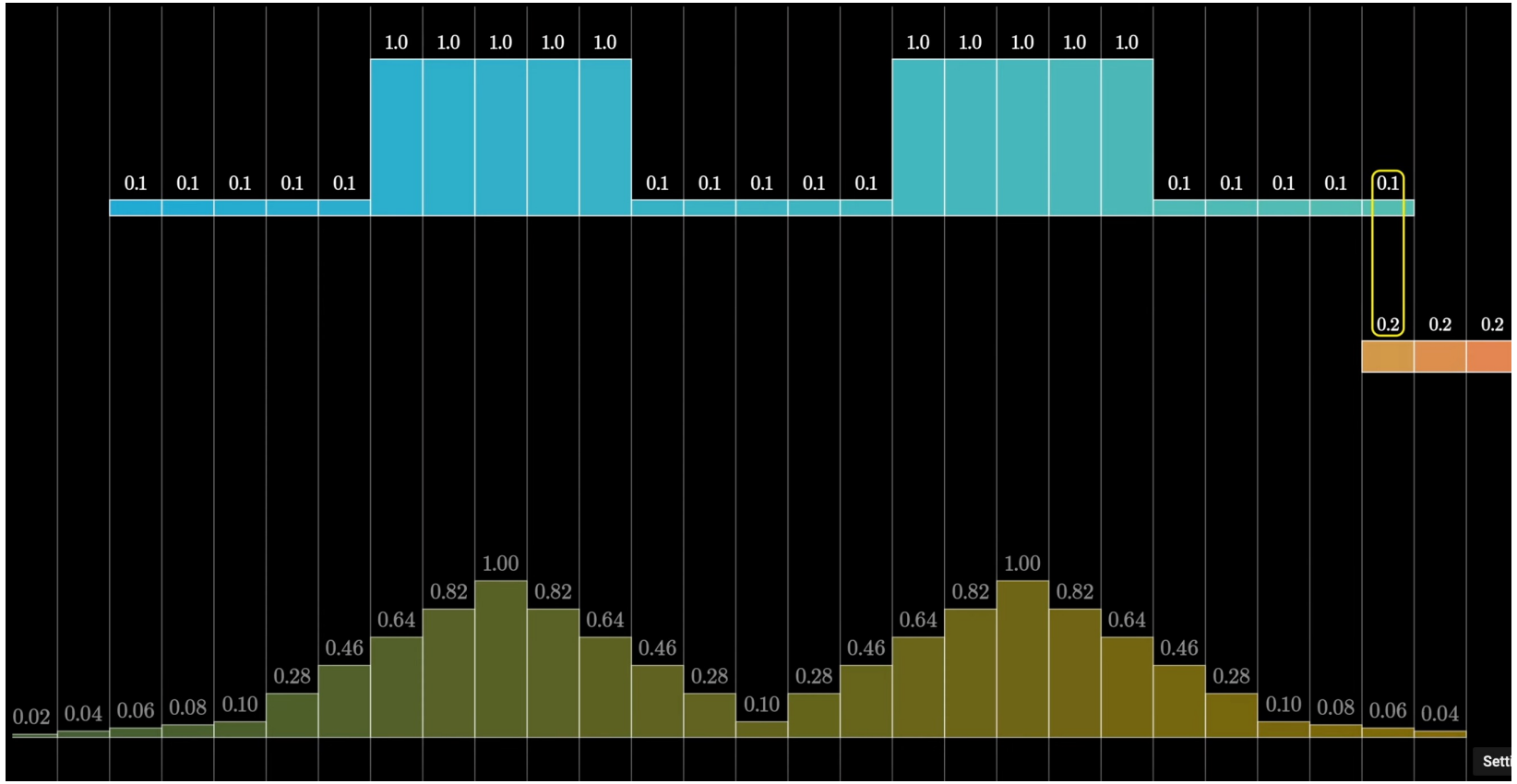
https://youtu.be/KuXjwB4LzSA

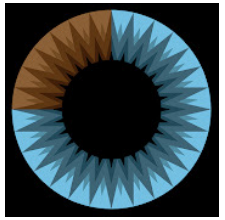
What kind of operation?



https://youtu.be/KuXjwB4LzSA

What kind of operation?

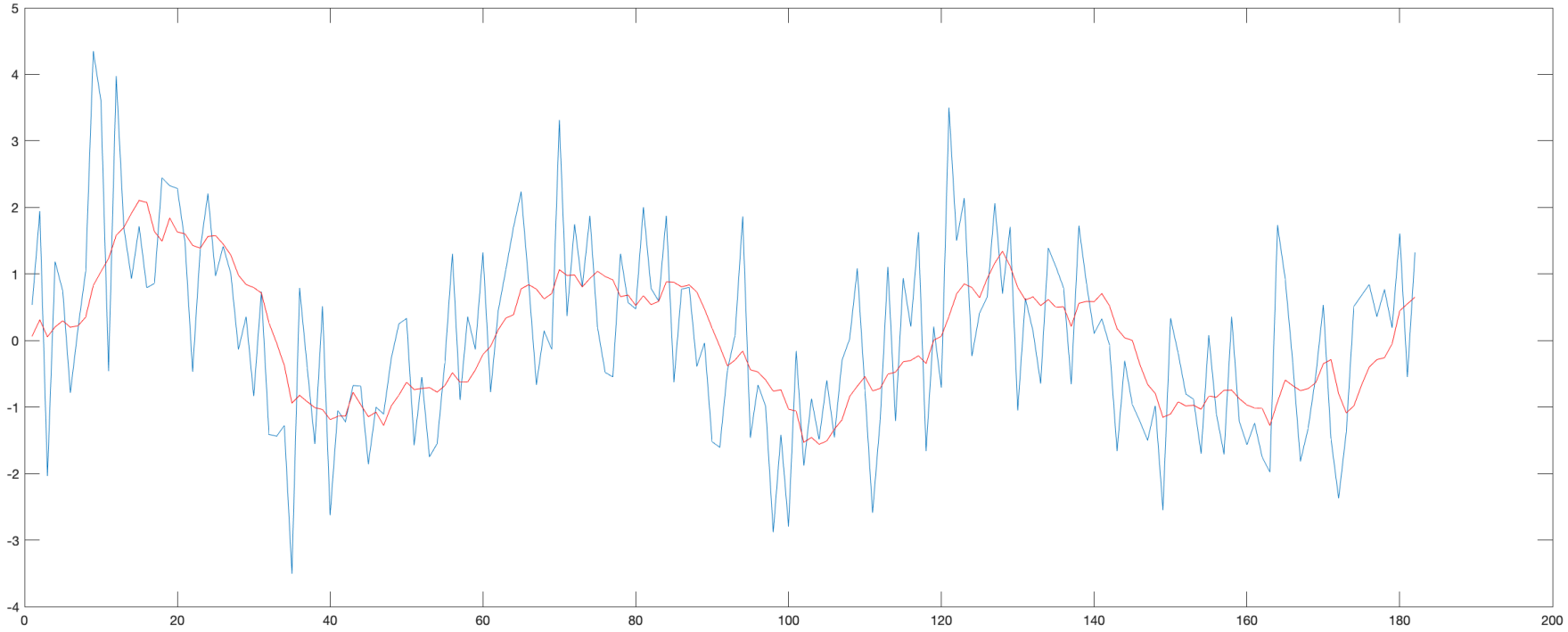




3Blue1Brown

https://youtu.be/KuXjwB4LzSA

Moving average – 1D Convolution



filter = 1/8 * [1 1 1 1 1 1 1 1];

2D Convolution

Input

7	2	3	3	8
4	5	3	8	4
3	3	2	8	4
2	8	7	2	7
5	4	4	5	4

*

Kernel/Filter

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

7x1+4x1+3x1+
2x0+5x0+3x0+
3x-1+3x-1+2x-1
= 6

=

Feature/Activation map

6		

https://medium.datadriveninvestor.com/convolutional-neural-networks-3b241a5da51e

Handcrafted filter for specific task

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

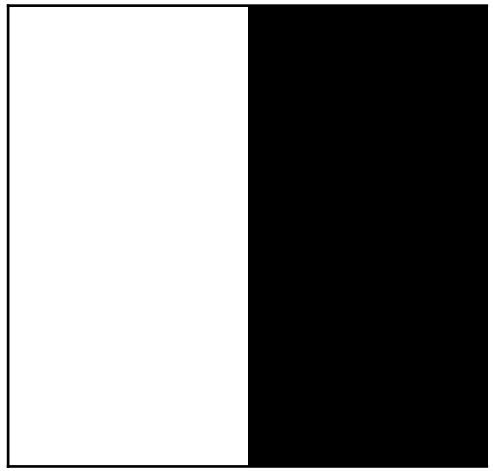
*

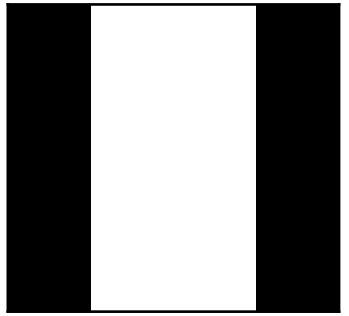
1	0	-1
1	0	-1
1	0	-1

Vertical edge detection filter

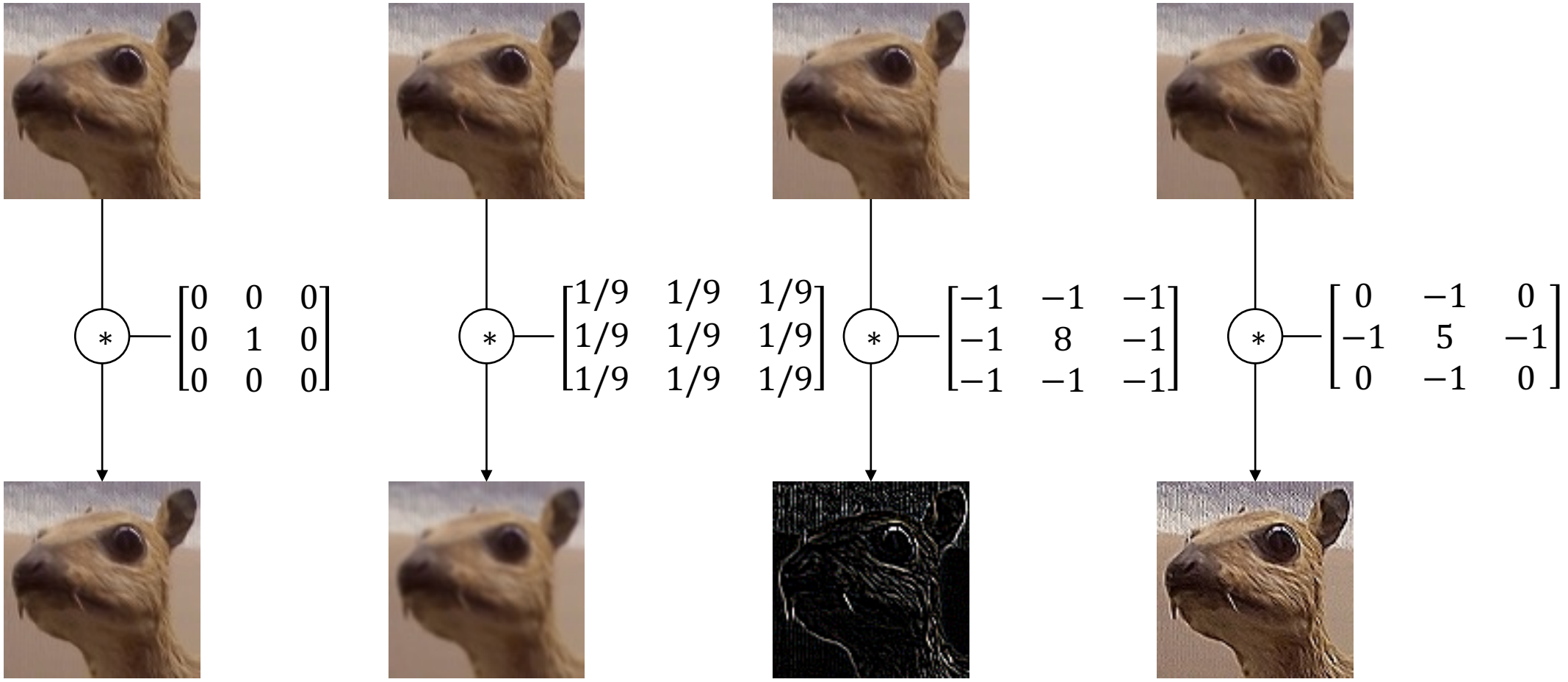
=

0	3	3	0
0	3	3	0
0	3	3	0
0	3	3	0



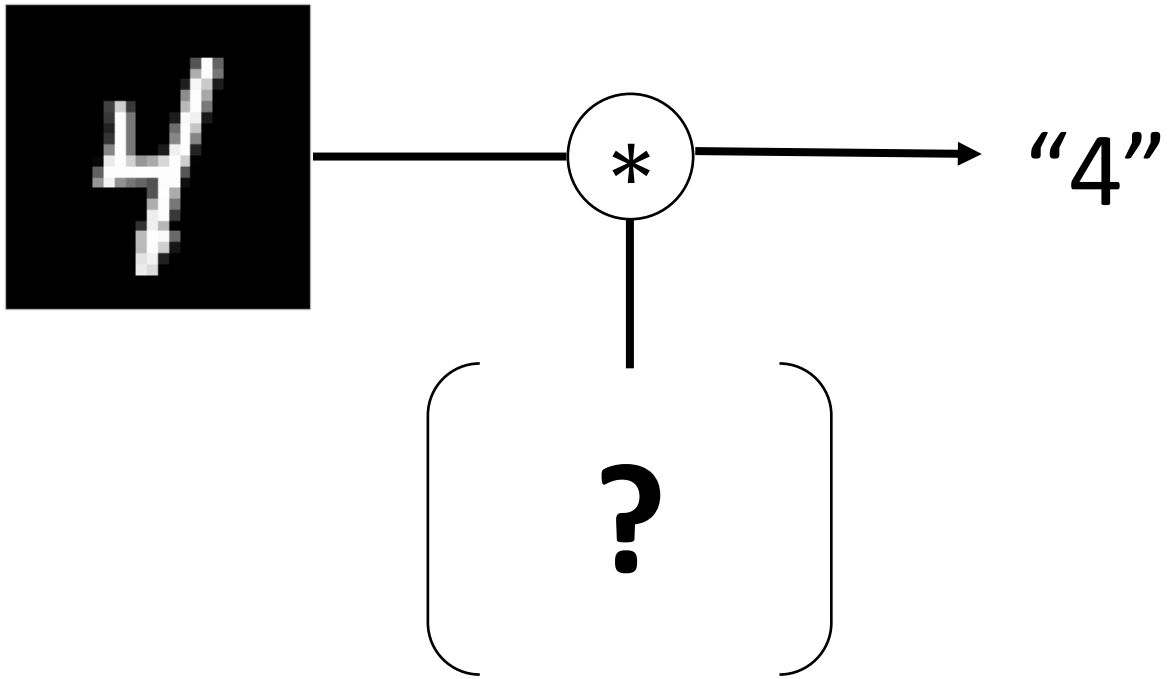


There are many specialized filters

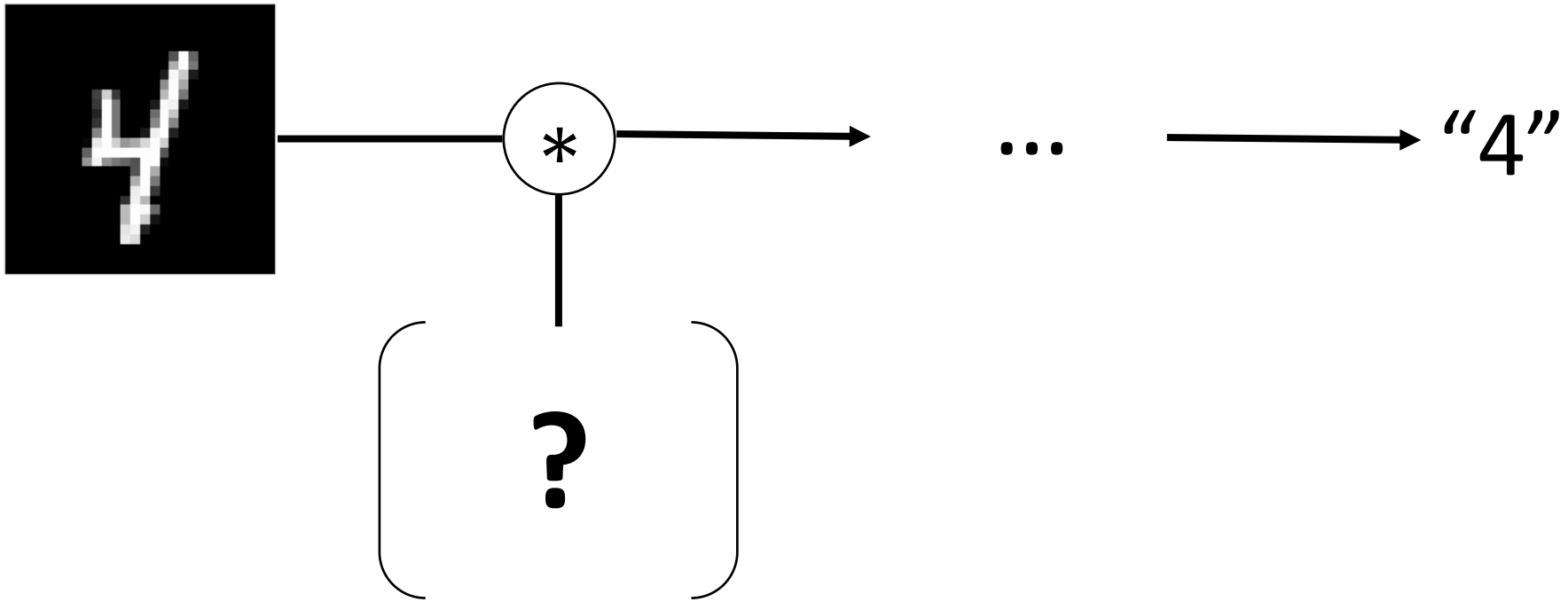


https://en.wikipedia.org/wiki/Kernel_(image_processing)

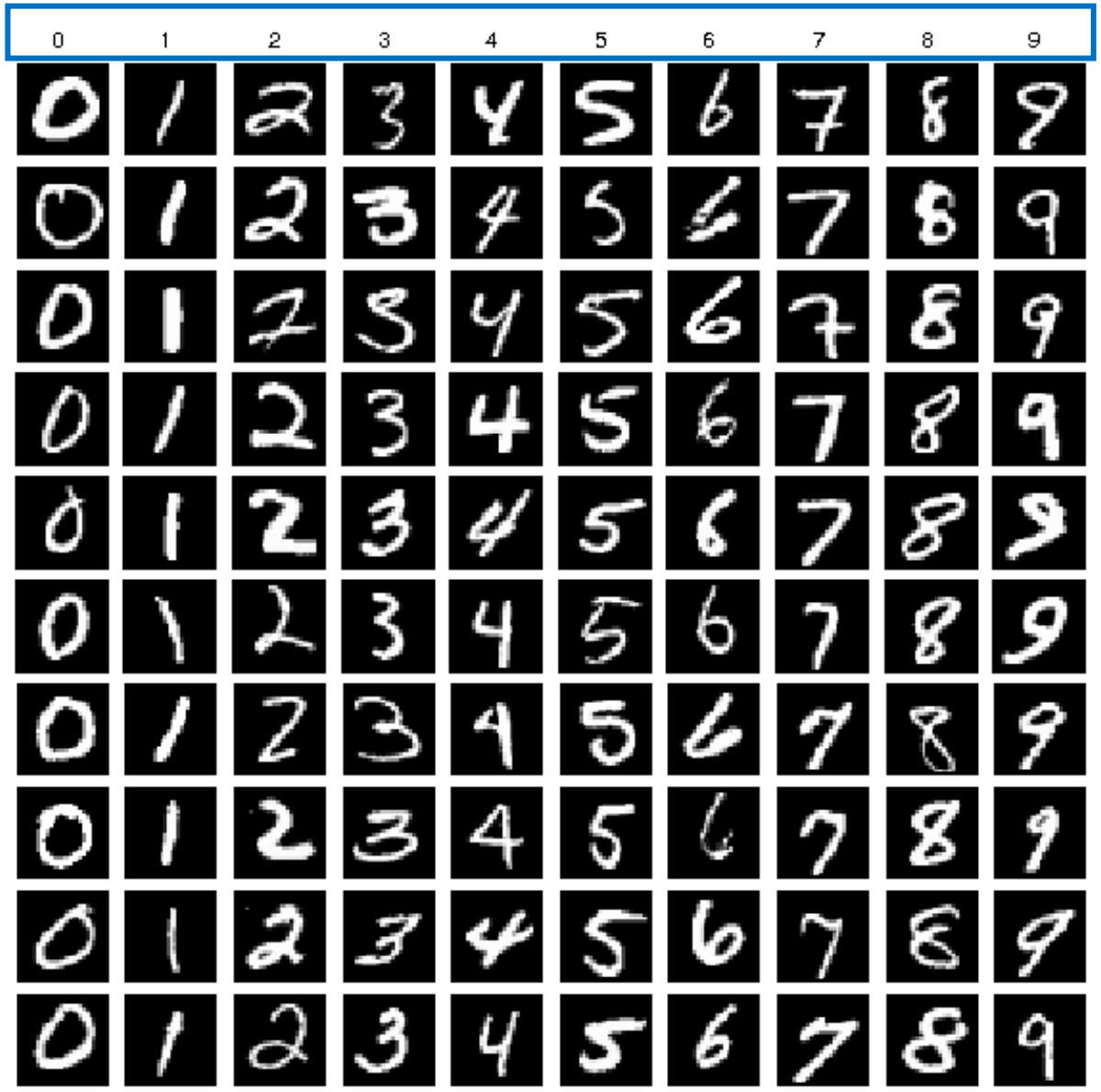
Filter for digit detection?



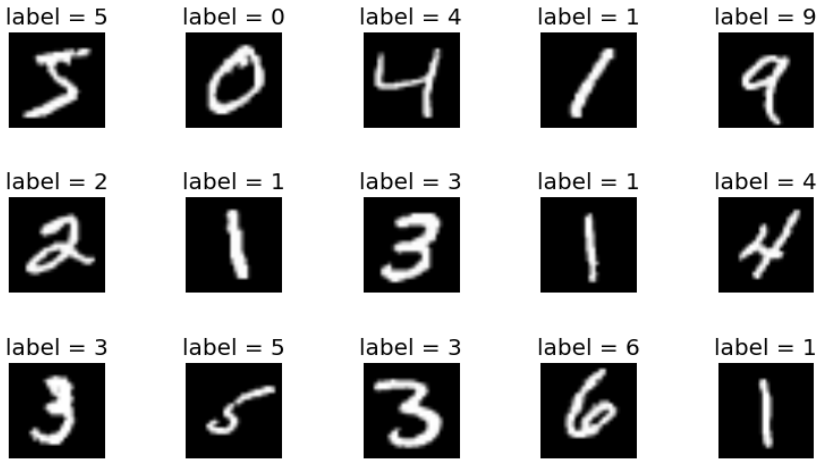
Filter for digit detection?



Let machine learn



Algorithm meets Data meets Compute



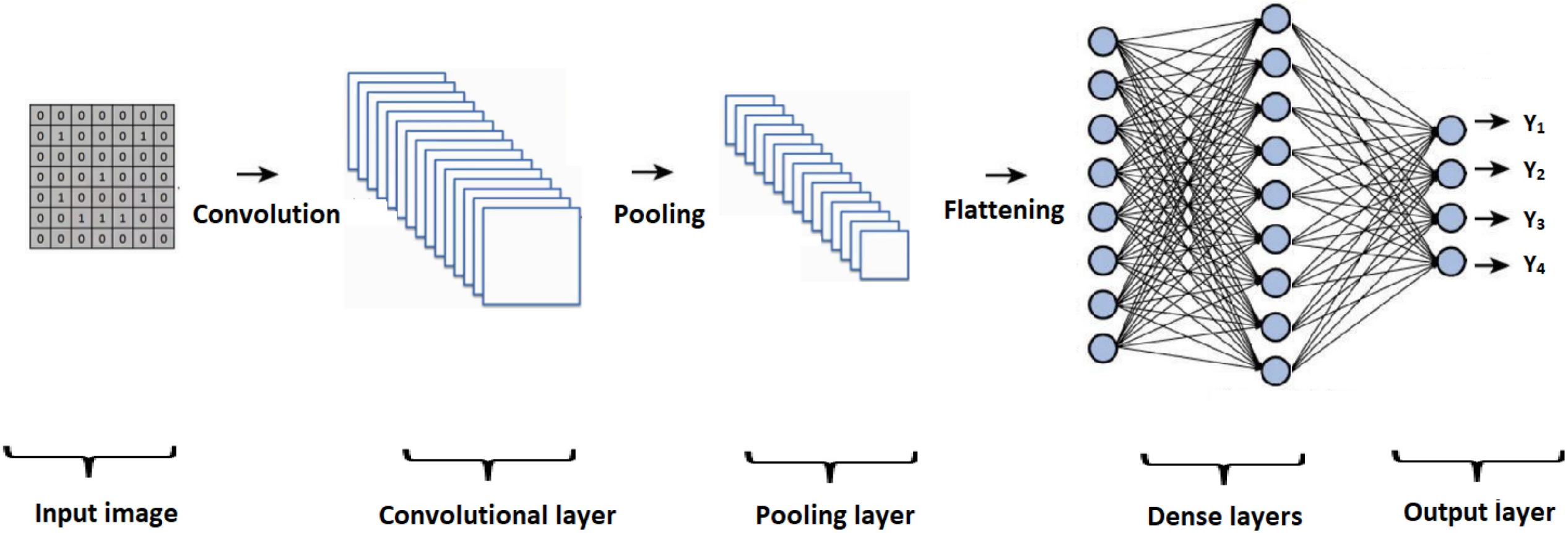
MNIST dataset

LeNet-5 (Convolutional Neural Network)

Digit

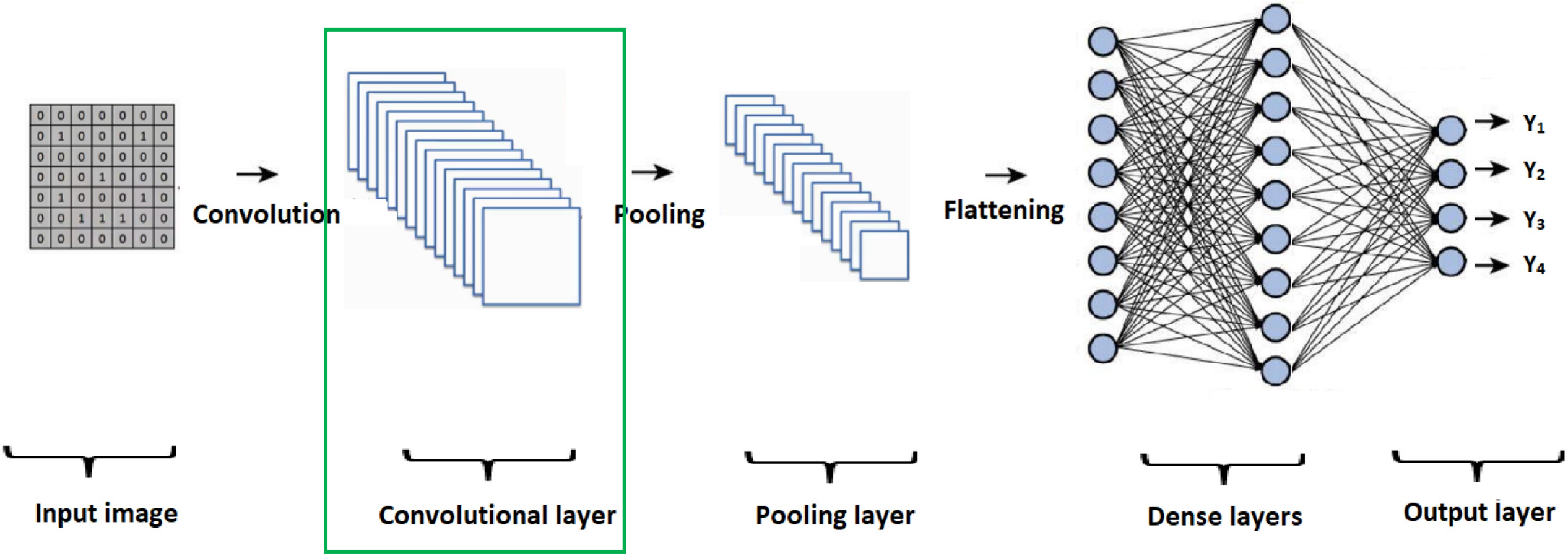
Y. Lecun, L. Bottou, Y. Bengio and P. Haffner, "Gradient-based learning applied to document recognition," in *Proceedings of the IEEE*, vol. 86, no. 11, pp. 2278-2324, Nov. 1998, doi: 10.1109/5.726791.

Convolutional Neural Network Architecture



https://www.mdpi.com/2076-3417/10/4/1245/htm

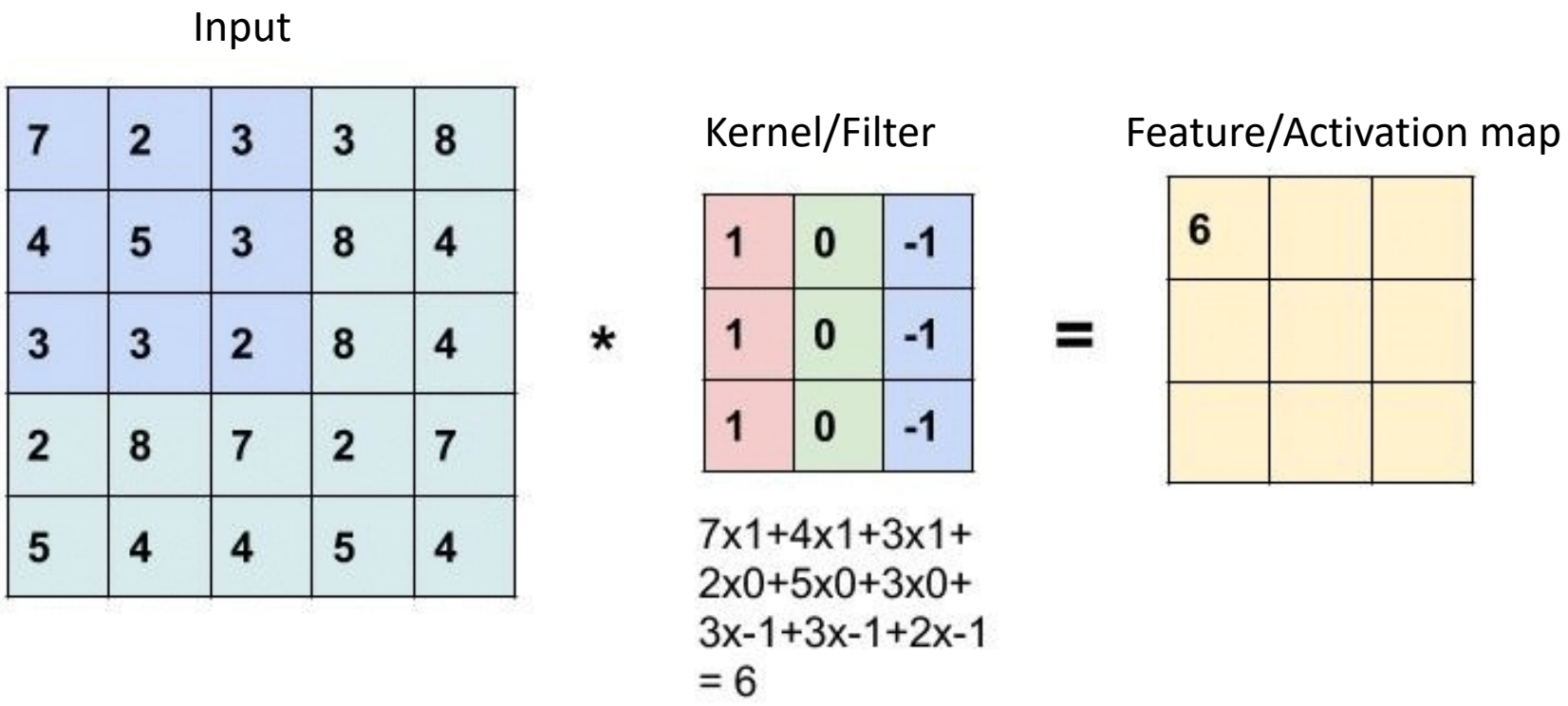
Convolutional Neural Network Architecture



https://www.mdpi.com/2076-3417/10/4/1245/htm

Convolutional layer

- Convolution operation



https://medium.datadriveninvestor.com/convolutional-neural-networks-3b241a5da51e

Convolutional layer

- Padding

0	0	0	0	0	0	0
0	60	113	56	139	85	0
0	73	121	54	84	128	0
0	131	99	70	129	127	0
0	80	57	115	69	134	0
0	104	126	123	95	130	0
0	0	0	0	0	0	0

Kernel

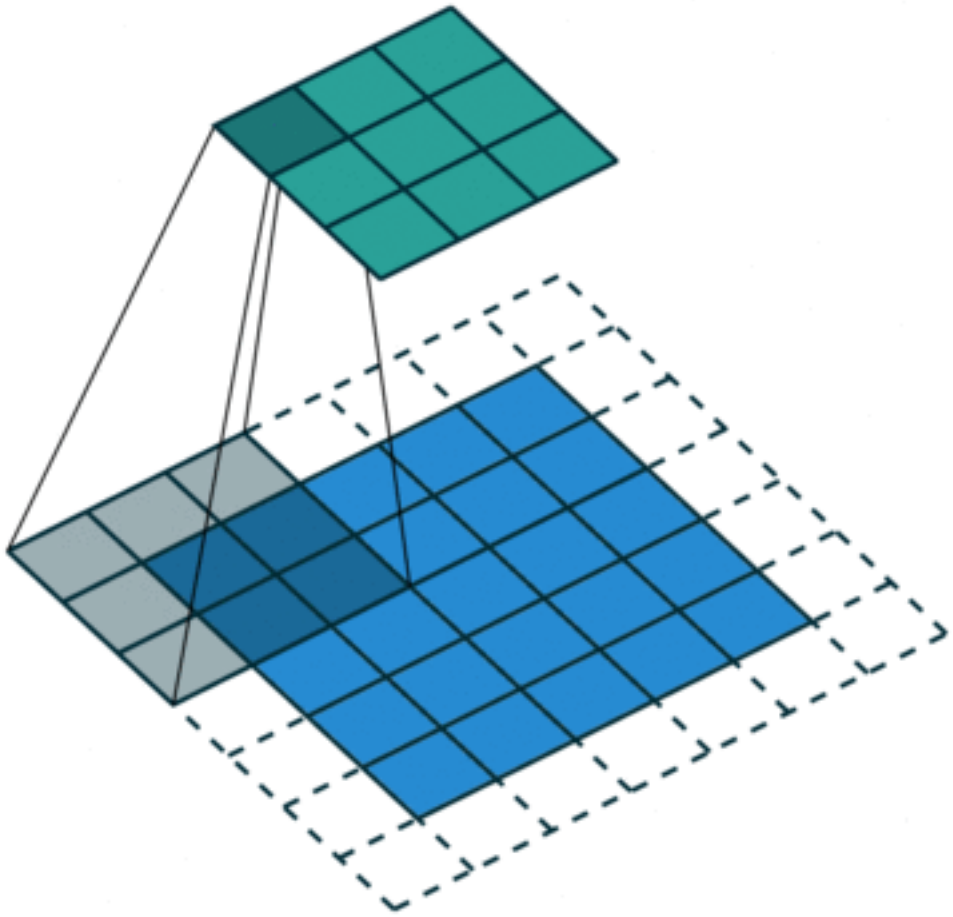
0	-1	0
-1	5	-1
0	-1	0

114				

https://www.pyimagesearch.com/2018/12/31/keras-conv2d-and-convolutional-layers/

Convolutional layer

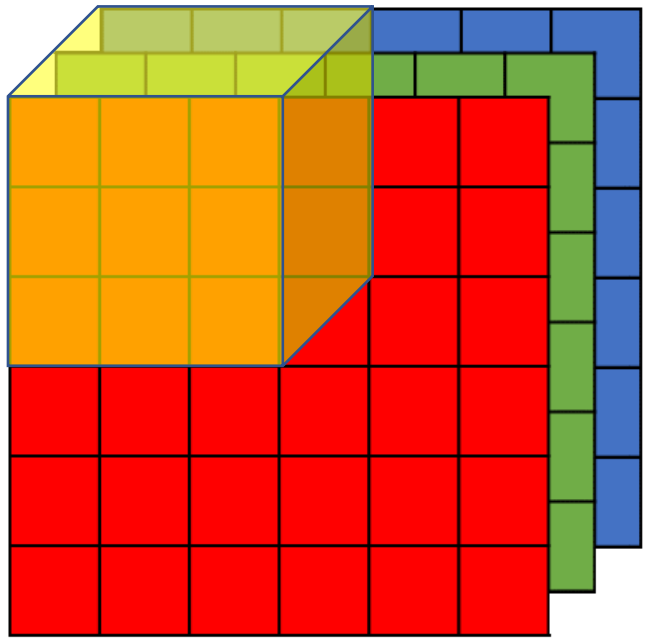
- Stride



Example: Padding = 1, Stride = 2

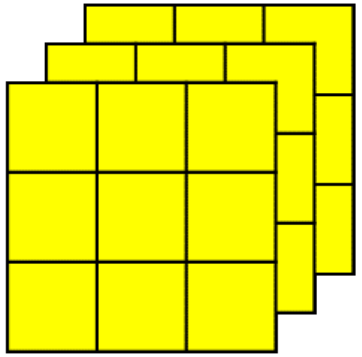
Convolutional layer

- Convolution over volume



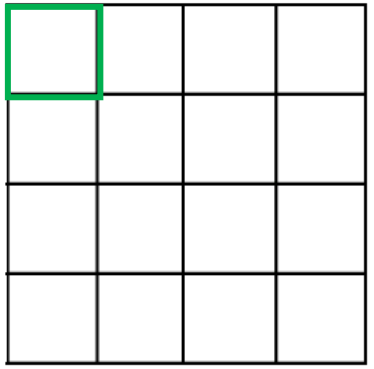
$6 \times 6 \times 3$

*



$3 \times 3 \times 3$

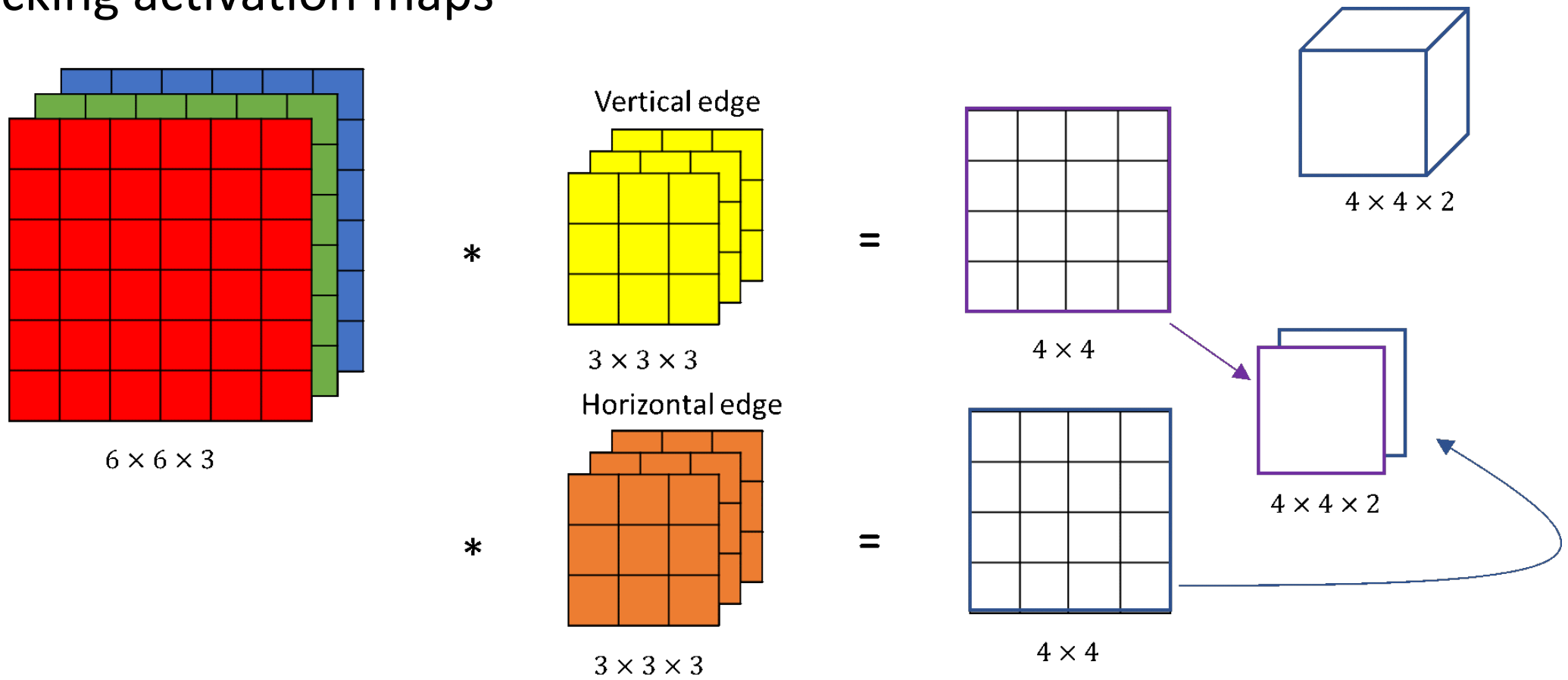
=



4×4

Convolutional layer

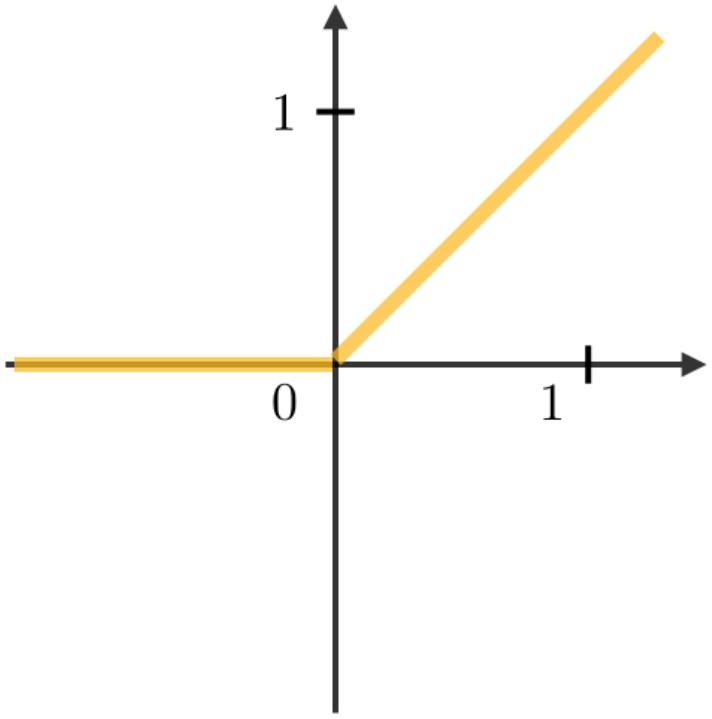
- Stacking activation maps



Activation functions

- Rectified Linear Unit (ReLU)

$$g(z) = \max(0, z)$$



Filter 1 Feature Map

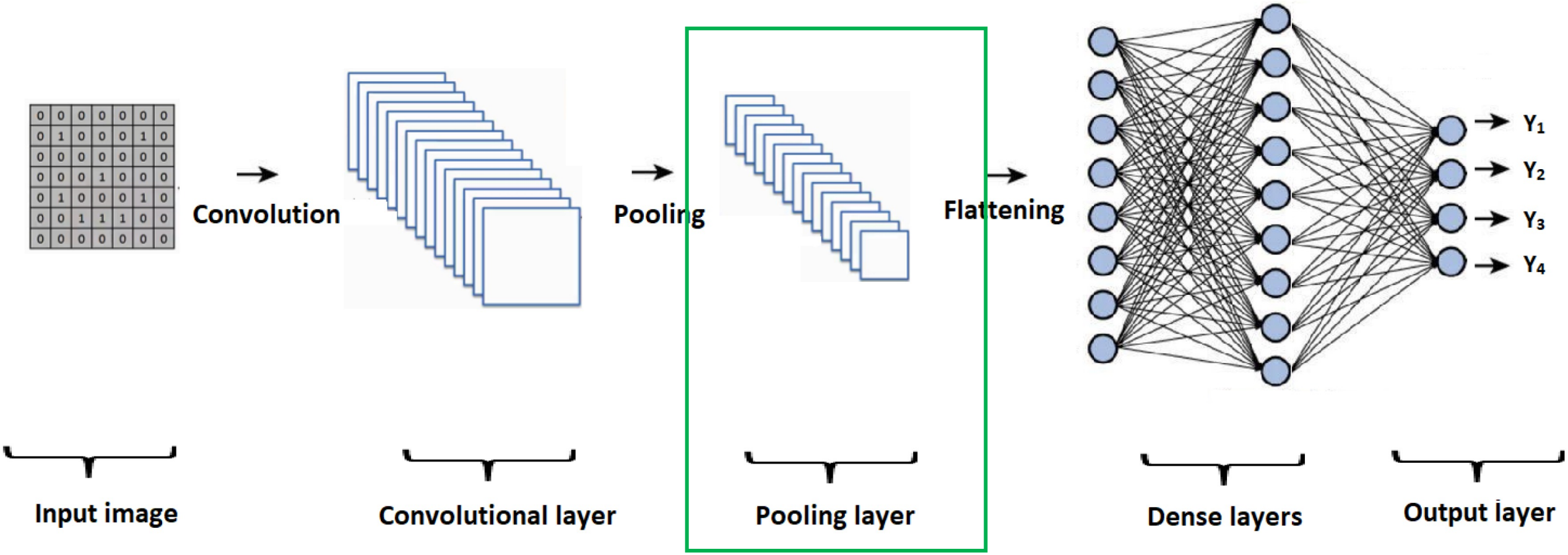
9	3	5	-8
-6	2	-3	1
1	3	4	1
3	-4	5	1

ReLU Layer

9	3	5	0
0	2	0	1
1	3	4	1
3	0	5	1

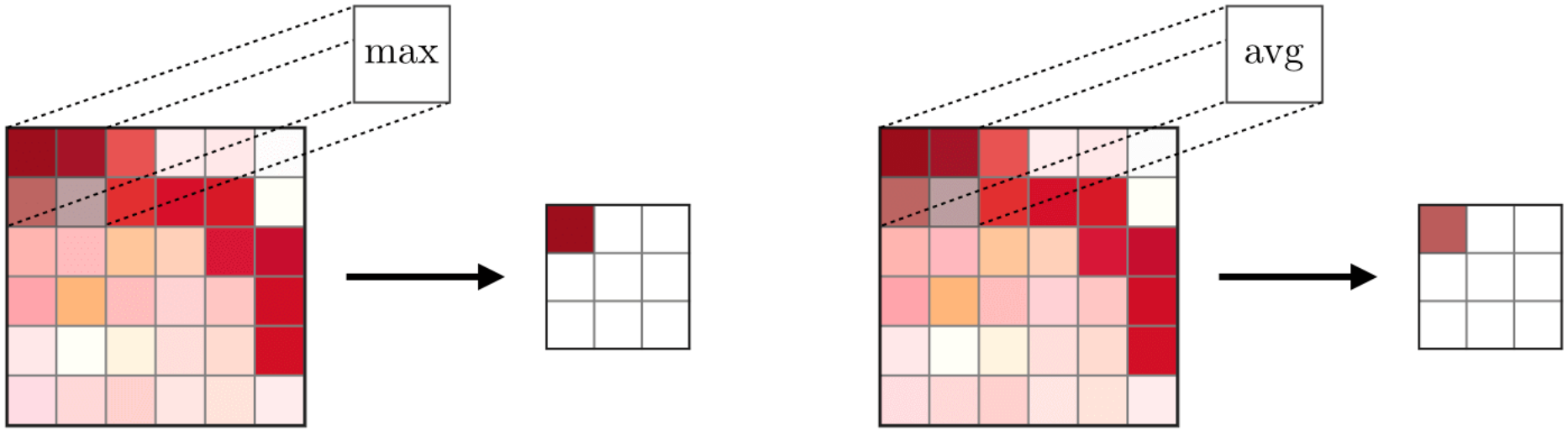
https://stanford.edu/~shervine/teaching/cs-230/cheatsheet-convolutional-neural-networks#layer

Convolutional Neural Network Architecture



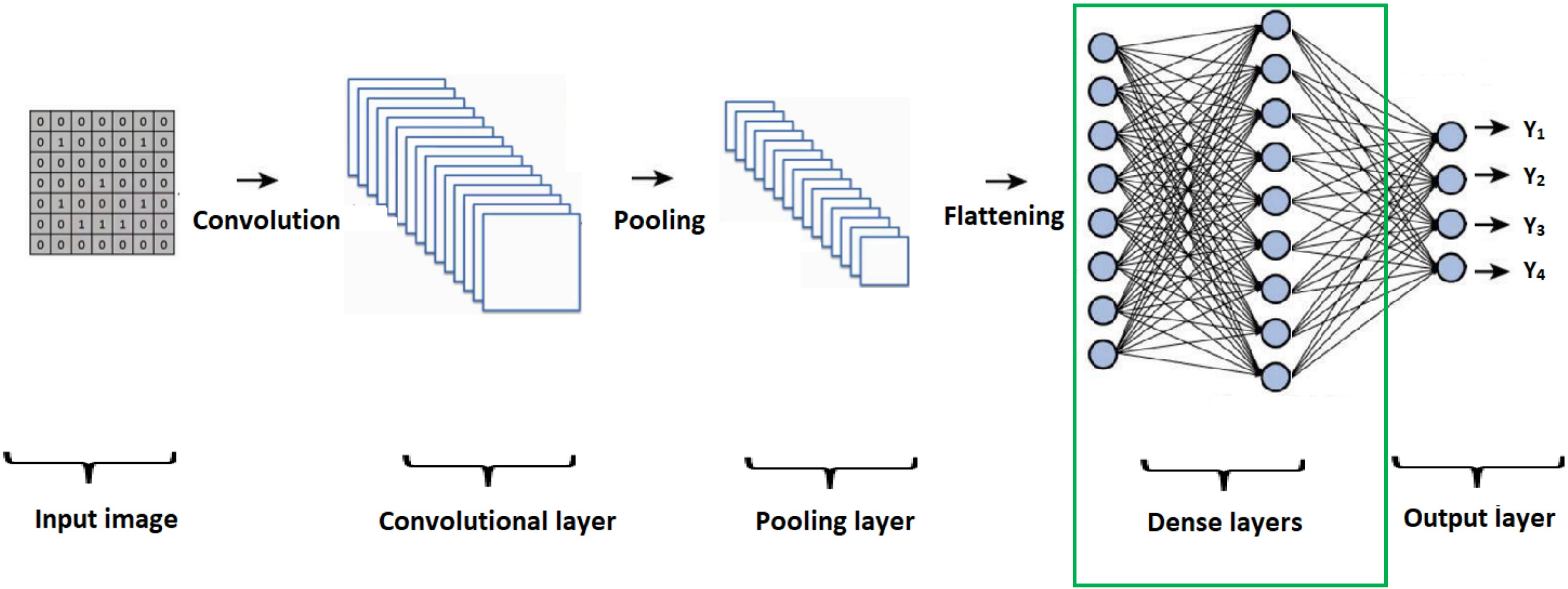
https://www.mdpi.com/2076-3417/10/4/1245/htm

Pooling layer



https://stanford.edu/~shervine/teaching/cs-230/cheatsheet-convolutional-neural-networks#layer

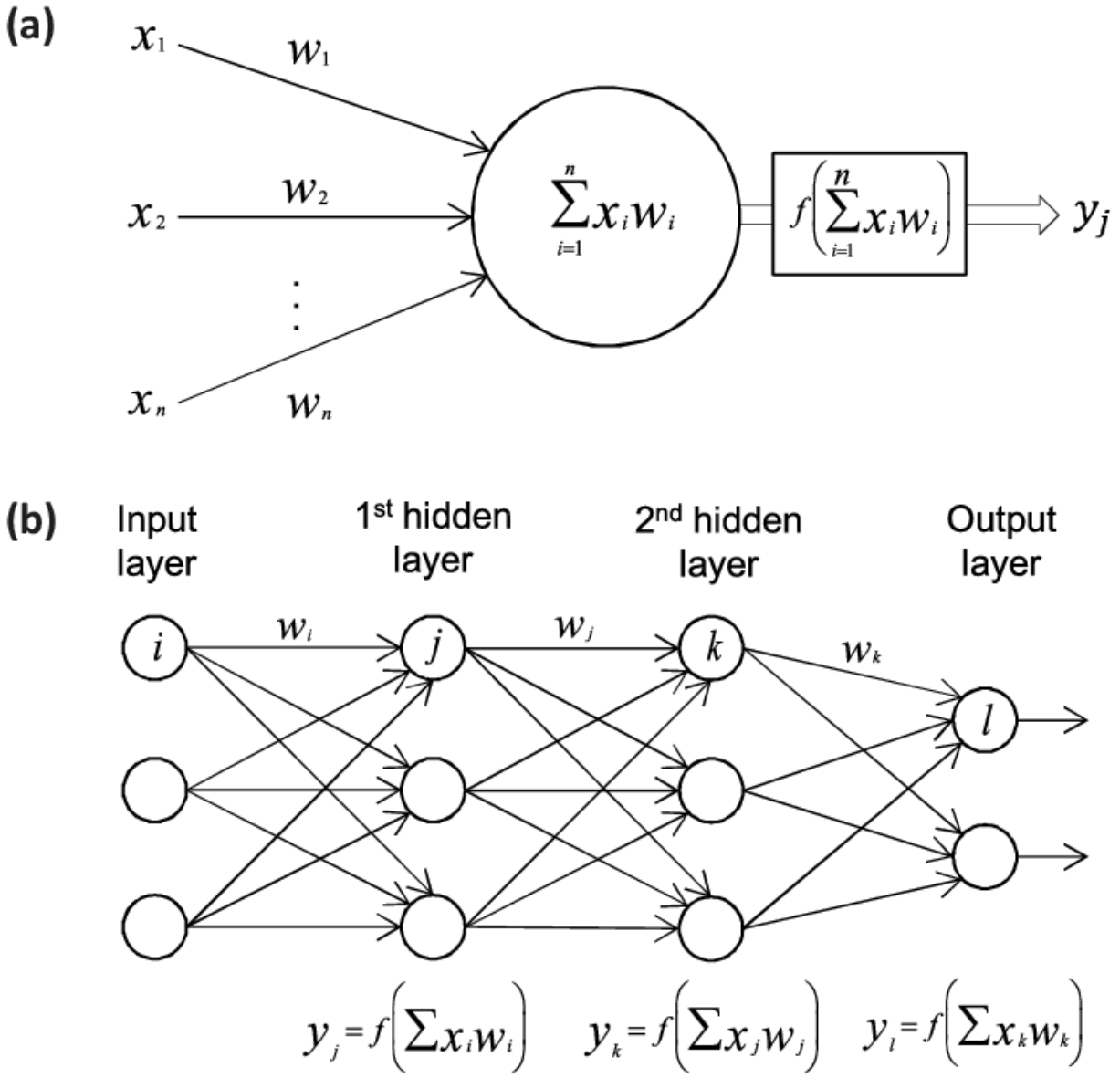
Convolutional Neural Network Architecture



https://www.mdpi.com/2076-3417/10/4/1245/htm

Fully connected layer

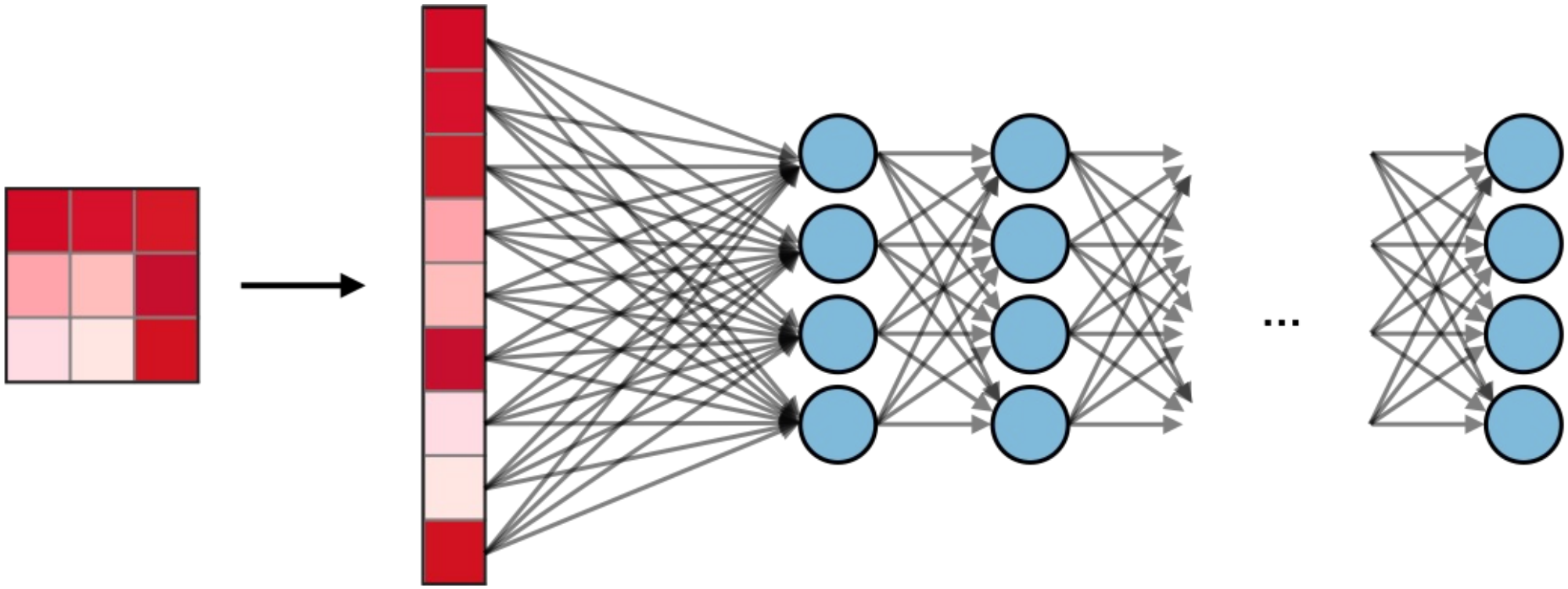
- Also known as dense layer
- Each input is connected to all hidden units



https://www.researchgate.net/figure/a-The-building-block-of-deep-neural-networks-artificial-neuron-or-node-Each-input-x_fig1_312205163

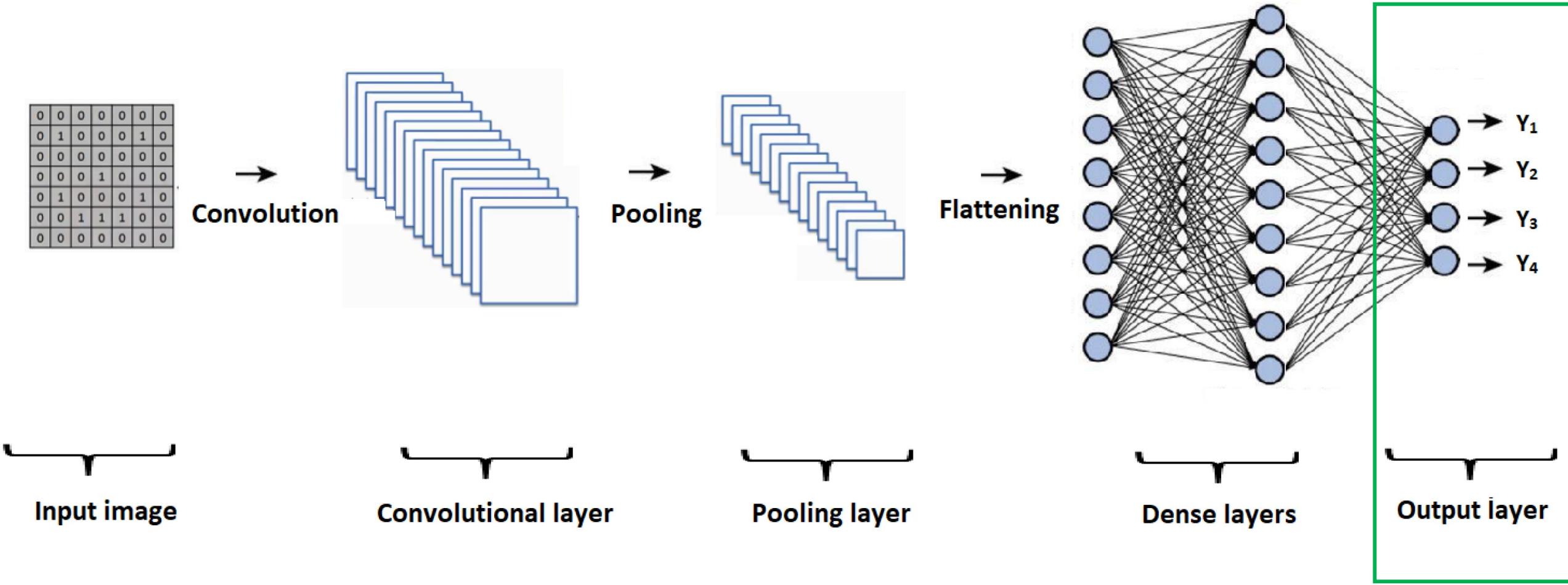
Fully connected layer

- Flattening



https://stanford.edu/~shervine/teaching/cs-230/cheatsheet-convolutional-neural-networks#layer

Convolutional Neural Network Architecture



https://www.mdpi.com/2076-3417/10/4/1245/htm

Activation function for classification

- Softmax

$$p = \begin{pmatrix} p_1 \\ \vdots \\ p_n \end{pmatrix} \quad \text{where} \quad p_i = \frac{e^{x_i}}{\sum_{j=1}^{n} e^{x_j}}$$

Loss function

- Cross-entropy loss

$$L = -\frac{1}{m}\sum_{i=1}^{m} y_i \cdot \log(\hat{y}_i)$$

y is a one-hot vector ([0,0,1,....,0])

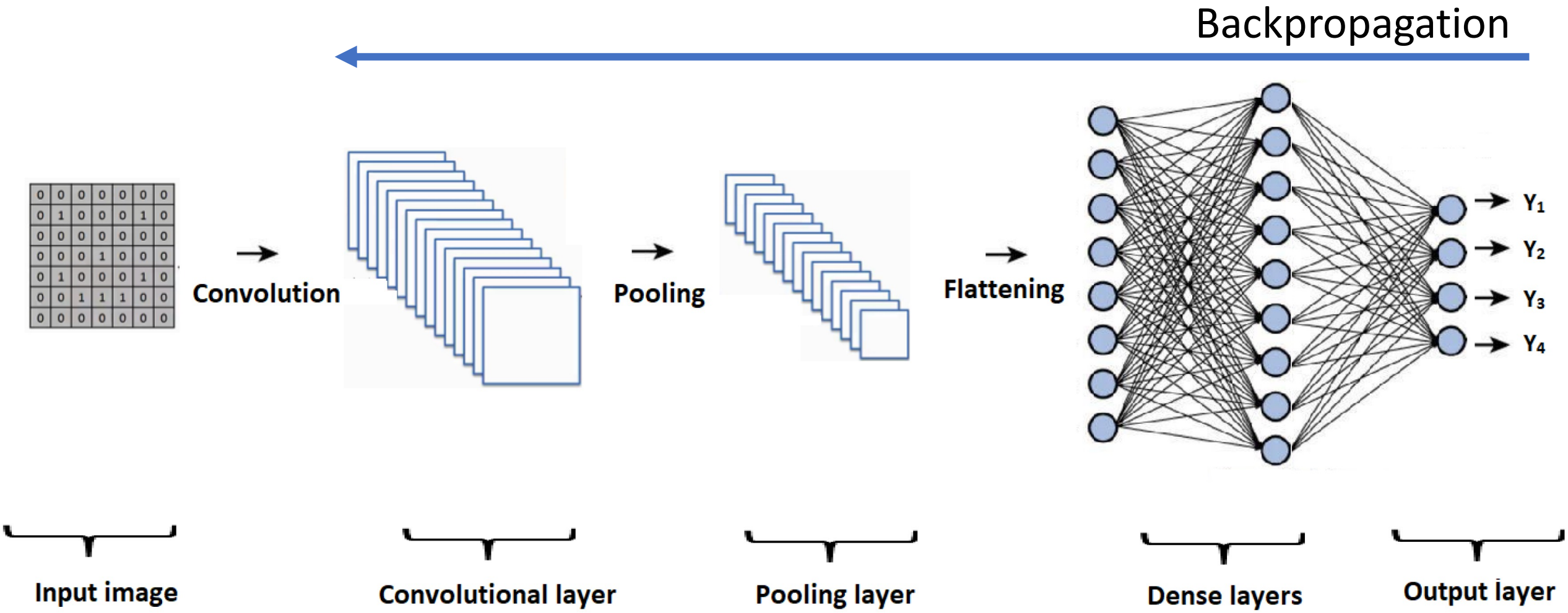
Gradient descent

$$\theta_j = \theta_j - \alpha \frac{\partial}{\partial \theta_j} L$$

Parameters

Learning rate

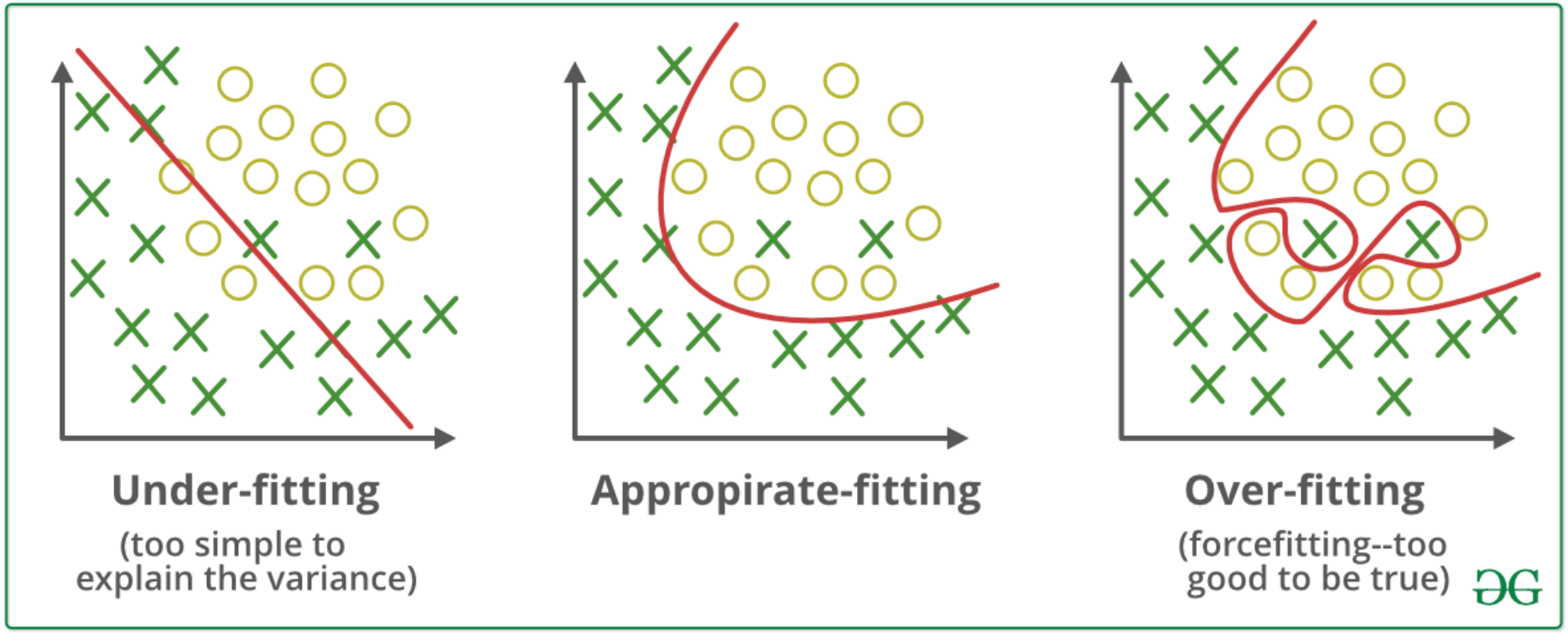
Training the network



https://www.mdpi.com/2076-3417/10/4/1245/htm

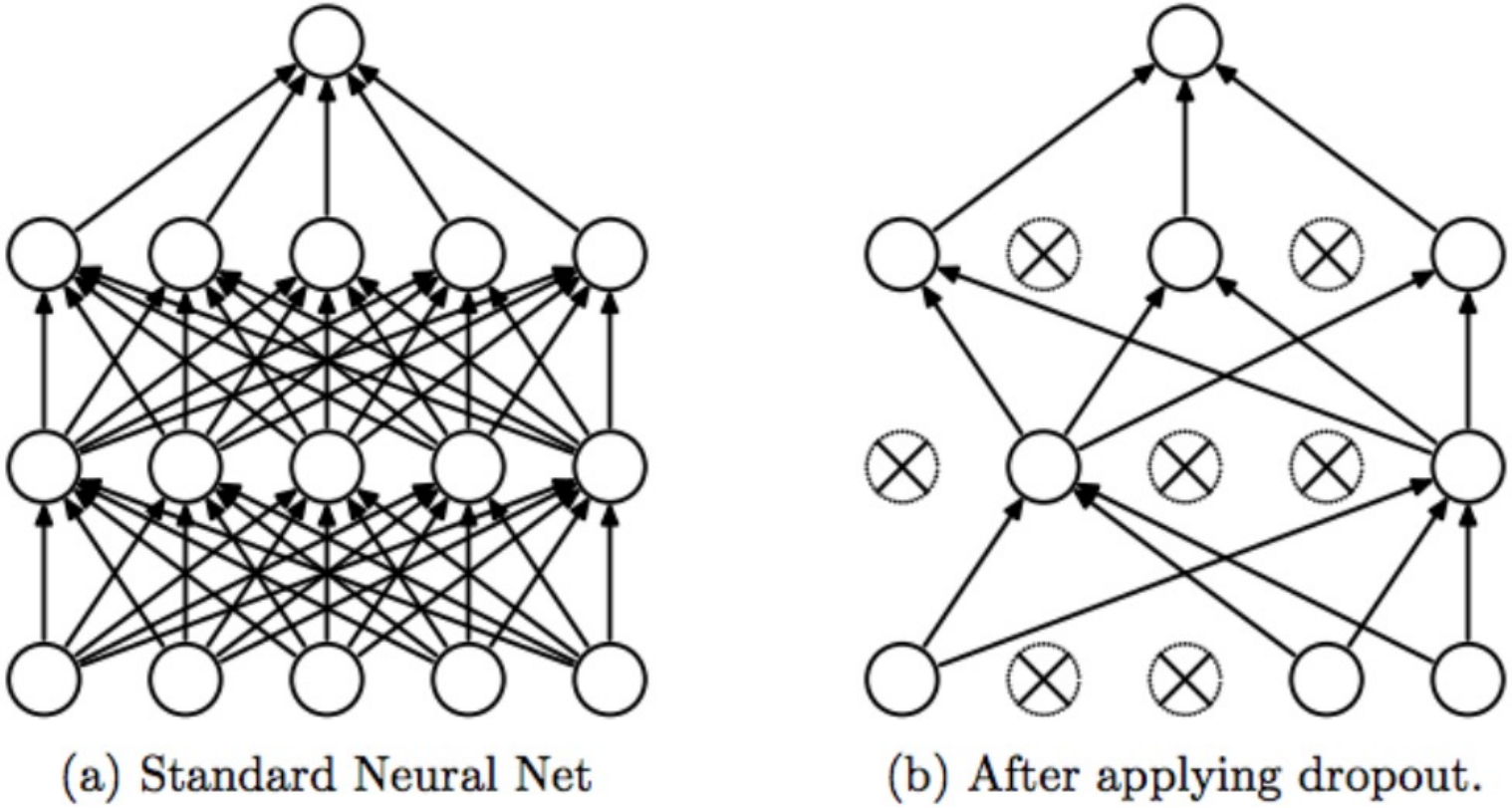
Dropout

- Overfitting



https://www.geeksforgeeks.org/underfitting-and-overfitting-in-machine-learning/

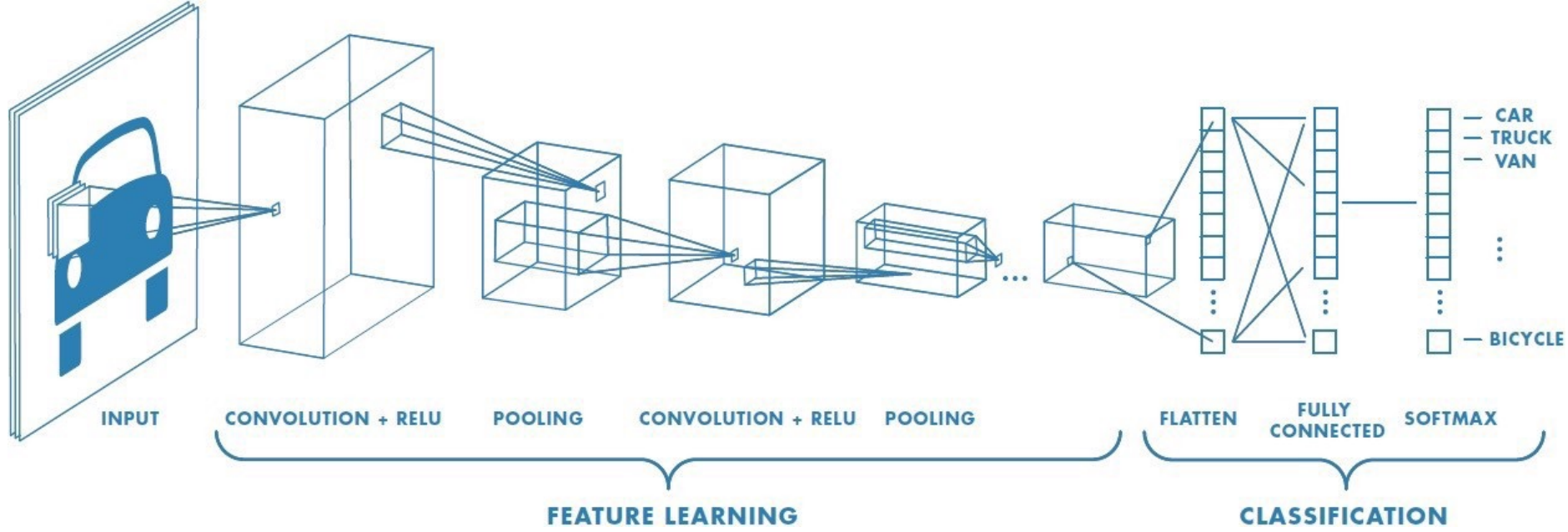
Dropout



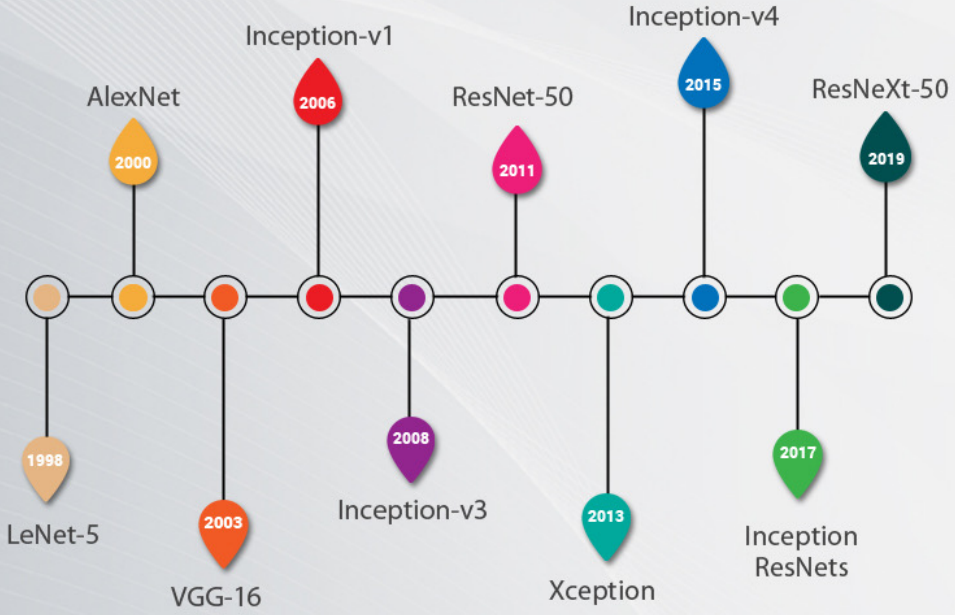
(a) Standard Neural Net

(b) After applying dropout.

https://medium.com/@amarbudhiraja/https-medium-com-amarbudhiraja-learning-less-to-learn-better-dropout-in-deep-machine-learning-74334da4bfc5



CNN architectures over a timeline(1998-2019)



IMDb

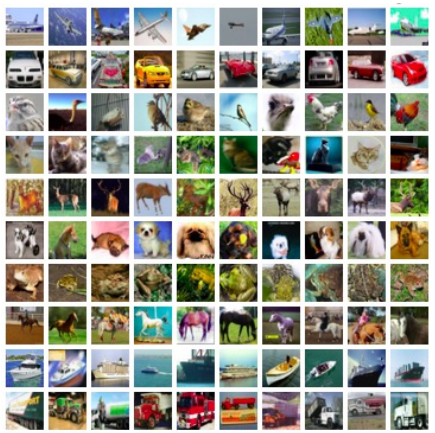
460,723 images

Wikipedia

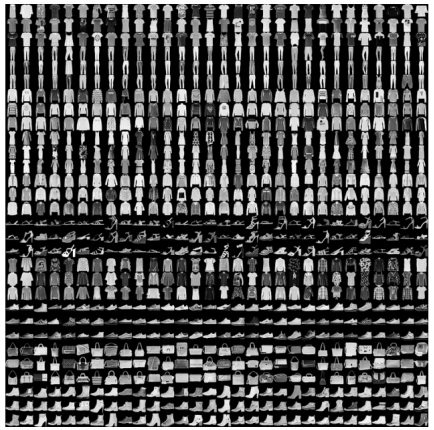
62,328 images

COCO 2020 Panoptic Segmentation Task

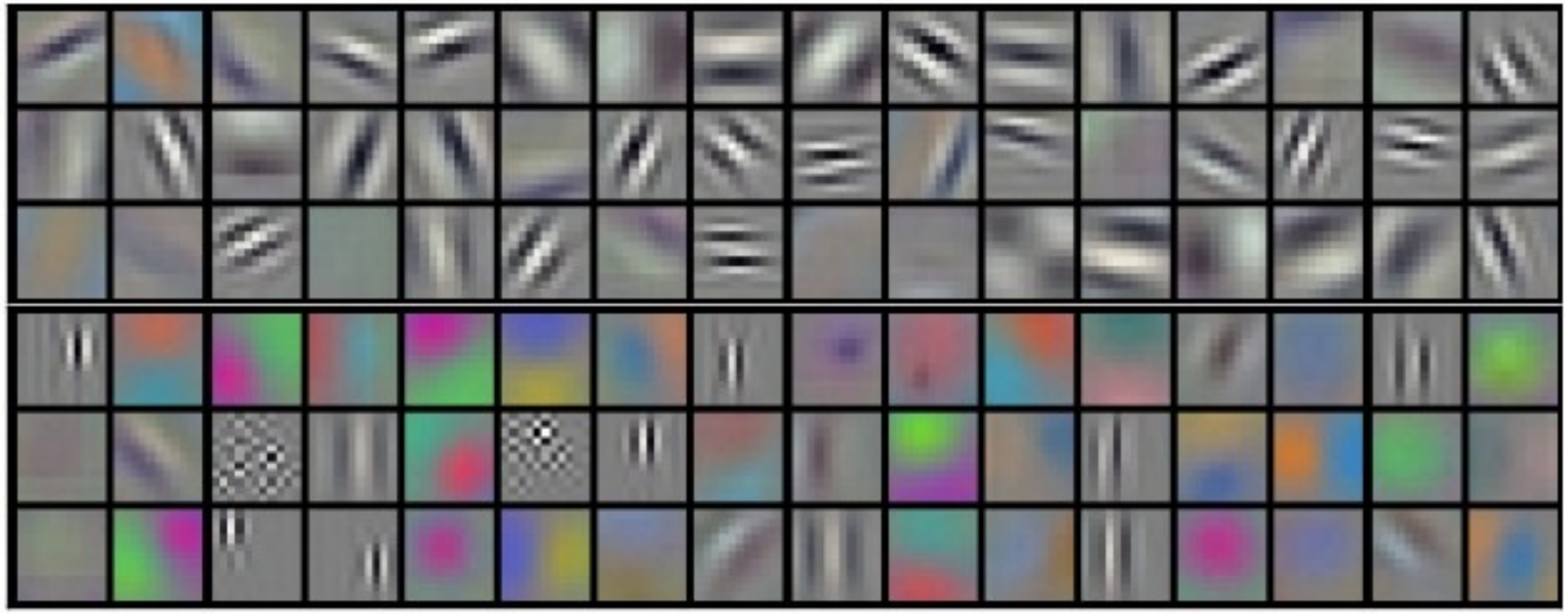
CIFAR-10



Fashion MNIST

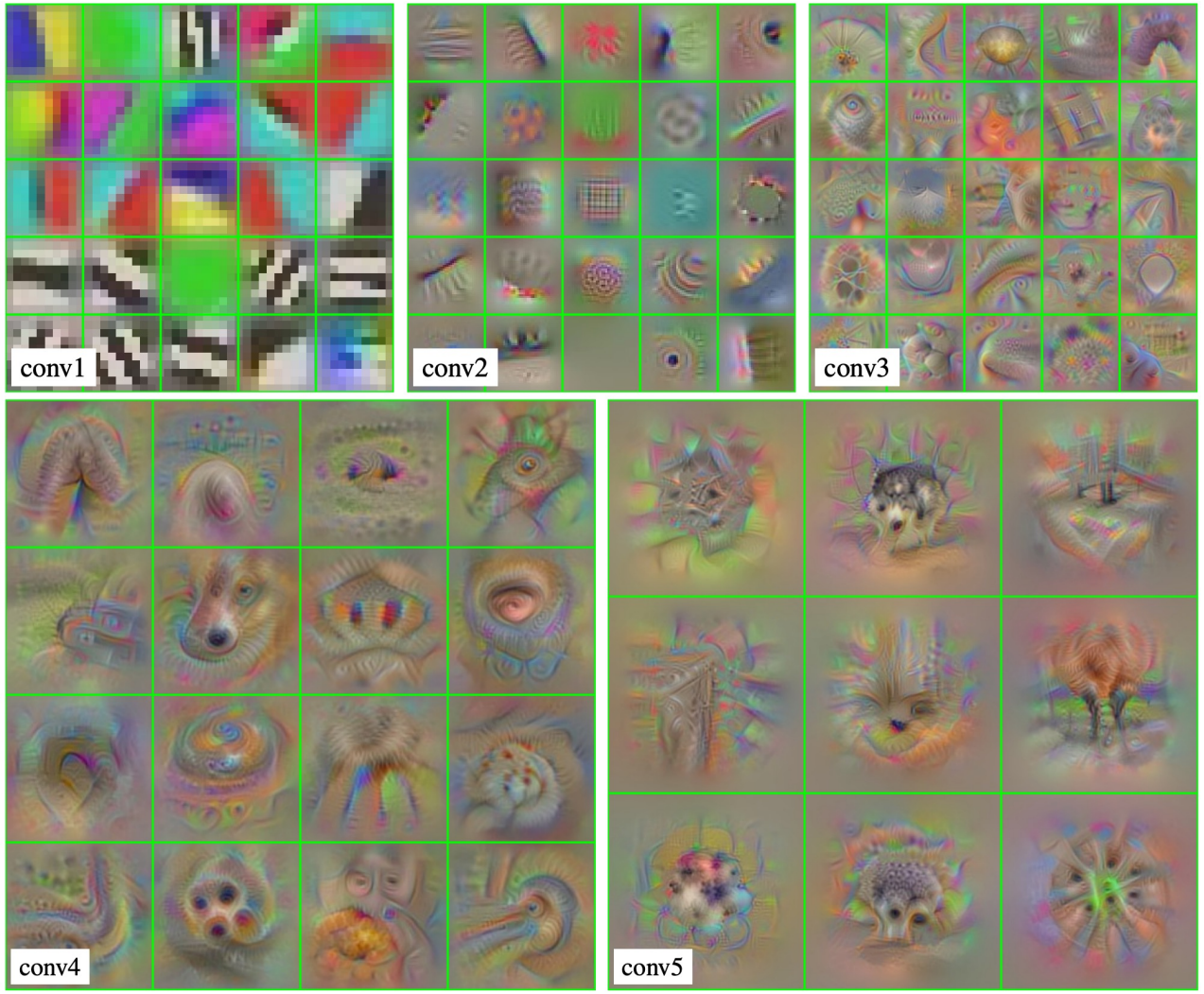


Visualizing learned filters



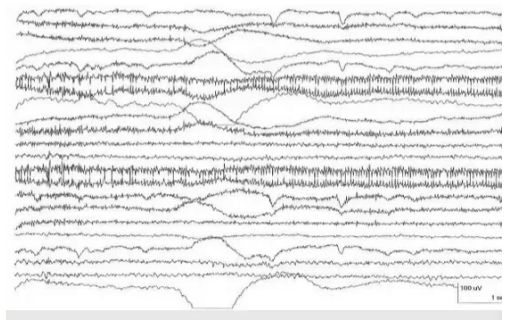
Krizhevsky et al., 2012

Visualizing learned filters

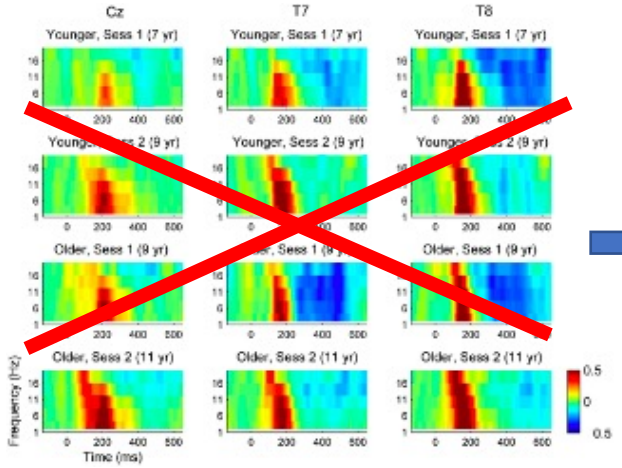


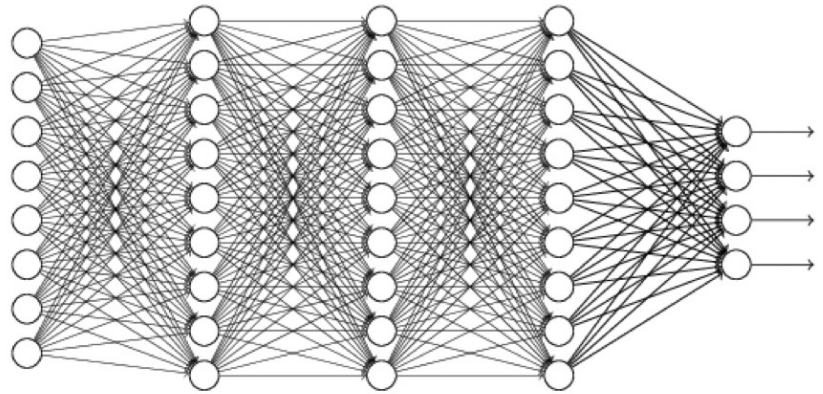
Mahendran & Vedaldi, 2016

Applying to EEG



ERSP





Emotion/Traits

Raw EEG vs. Frequency-domain

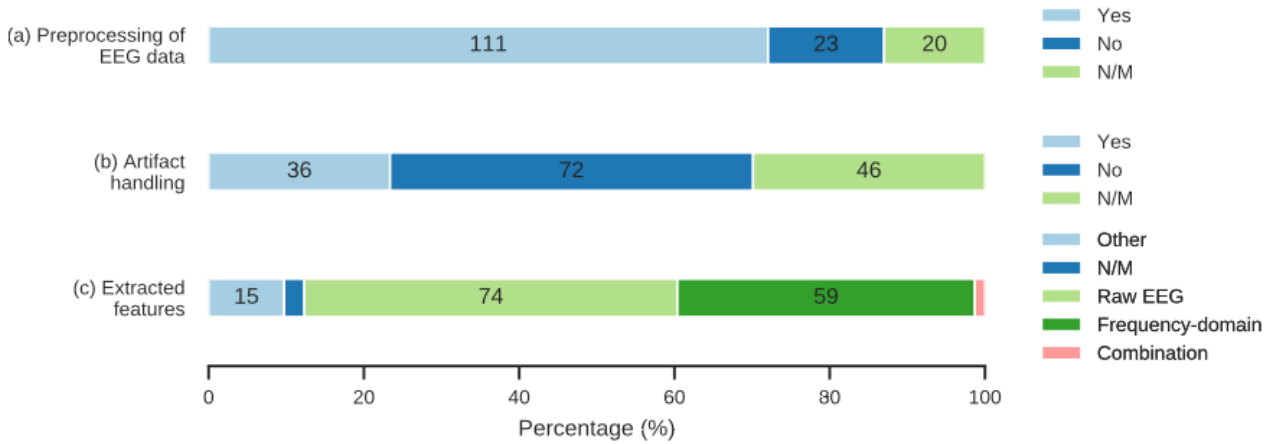
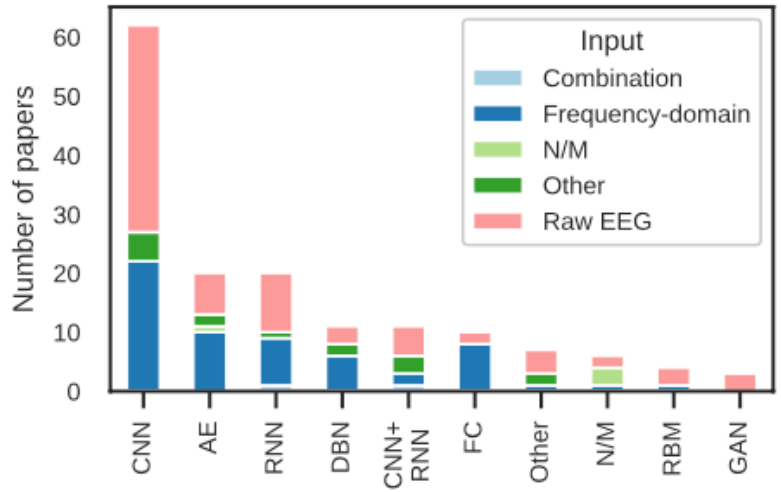


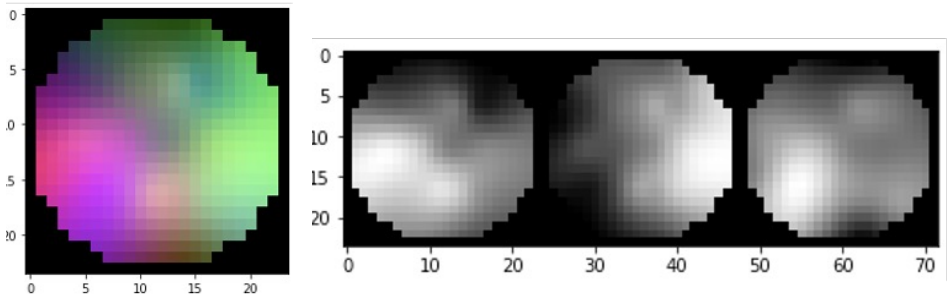
Figure 9. EEG processing choices. (a) Number of studies that used preprocessing steps, such as filtering, (b) number of studies that included, rejected or corrected artifacts in their data and (c) types of features that were used as input to the proposed models.



Distribution of input type according to the architecture category.

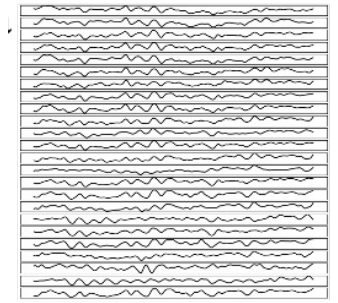
Roy Y. et al (2019)

Our approach

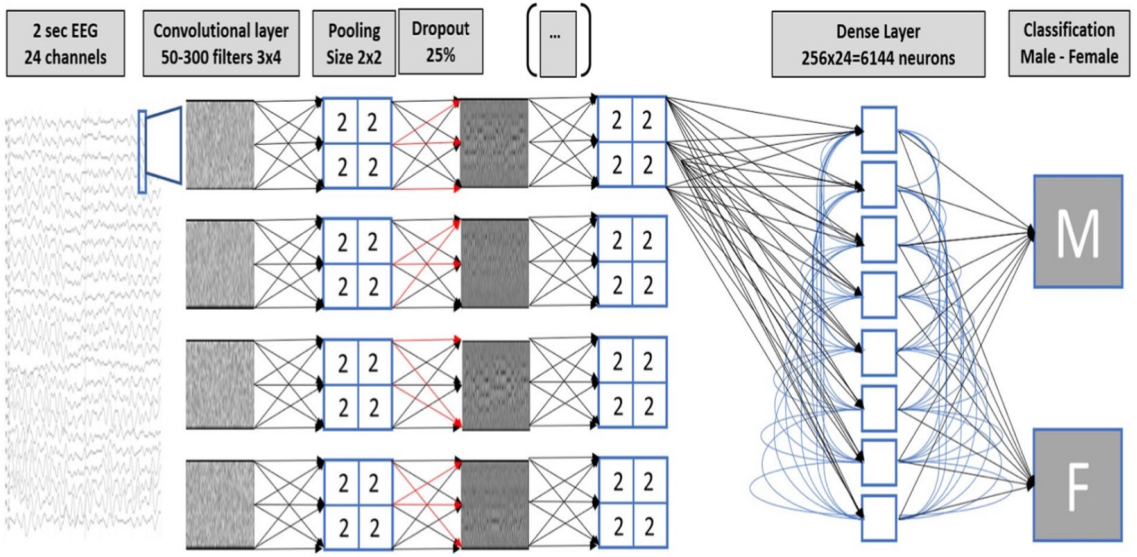


Spectral features

vs.

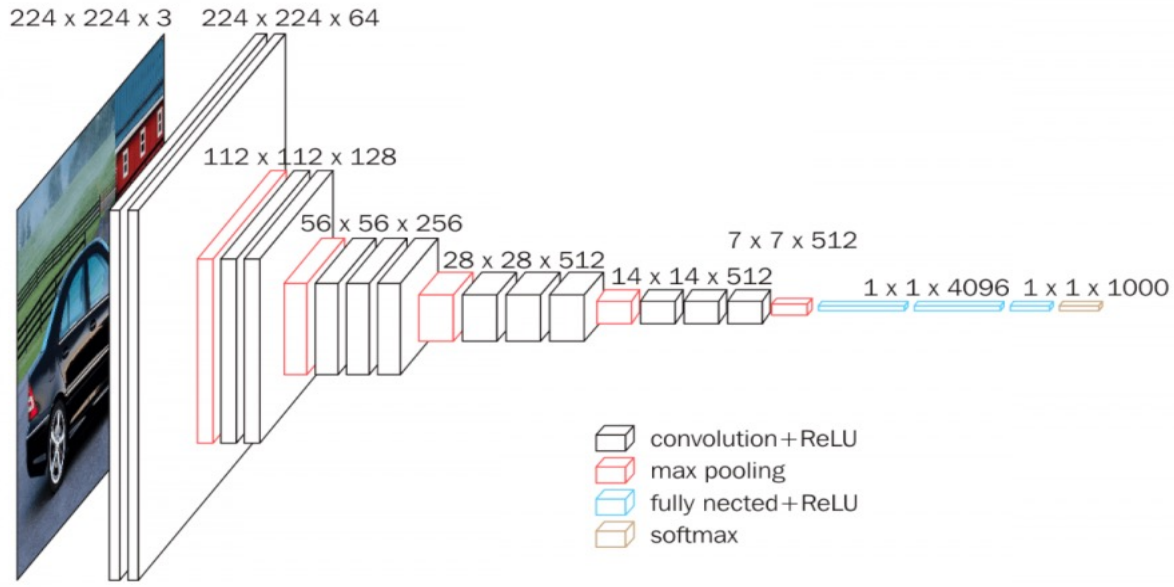


Raw EEG



van Putten et al. (2018)

vs.



Simonyan, K. and Zisserman A. (2014)

Data Data Data Data Data Data Data Data

CHILD MIND® INSTITUTE

Michael Milham, PI

- 128 channels EEG
- ~2,000 EEG datasets (planed 10,000)
- Tasks involving emotions (The Gift movie)
- Rest (eyes open and eyes closed)

Data

- 2224 subjects:
  - 787 female (35%)
  - Ages: min 5, max 22, mean 10
- Pre-processing following the paper:
  - Remove baseline
  - Filter 0.25-25Hz
  - Resample 128Hz
  - Re-reference to average mastoids
  - Epoching: eye-closed, 3 40-second blocks. Ignored first and last 3 seconds of each block
  - clean_rawdata – ASR (our)
  - Sub-select 24 channels
    - Fp1, Fp2, F7, F3, Fz, F4, F8, FC3, FCz, FC4, T3, C3, C4, T4, CP3, CPz, CP4, T5, P3, Pz, P4, T6, O1, Cz
  - Segment 2-second non-overlapping windows
    - → ~ 81 samples per subject

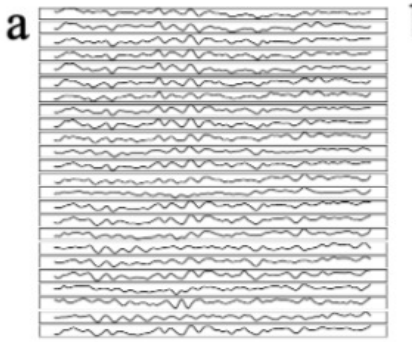
Data

- Sub-select 1574 participants (50% female) 24-channels
- 2-second extracted epochs eyes open and eyes closed
- 5 predictors: sex, handedness, eyes open/closed, age, segment count

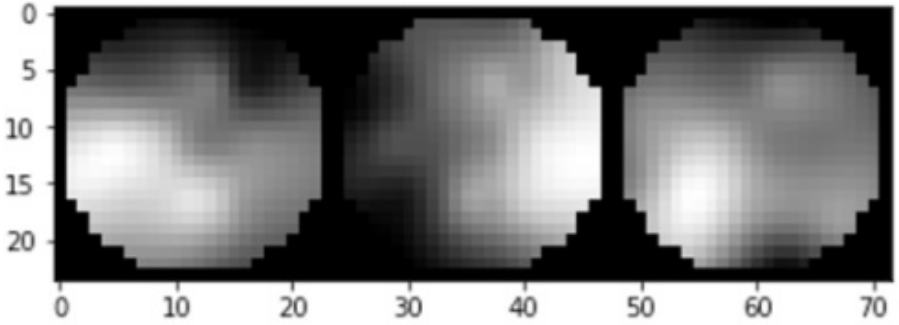
Categorical

Continuous

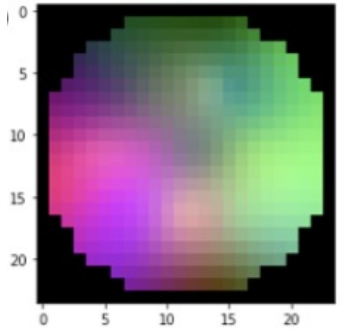
2-sec



Achromatic image

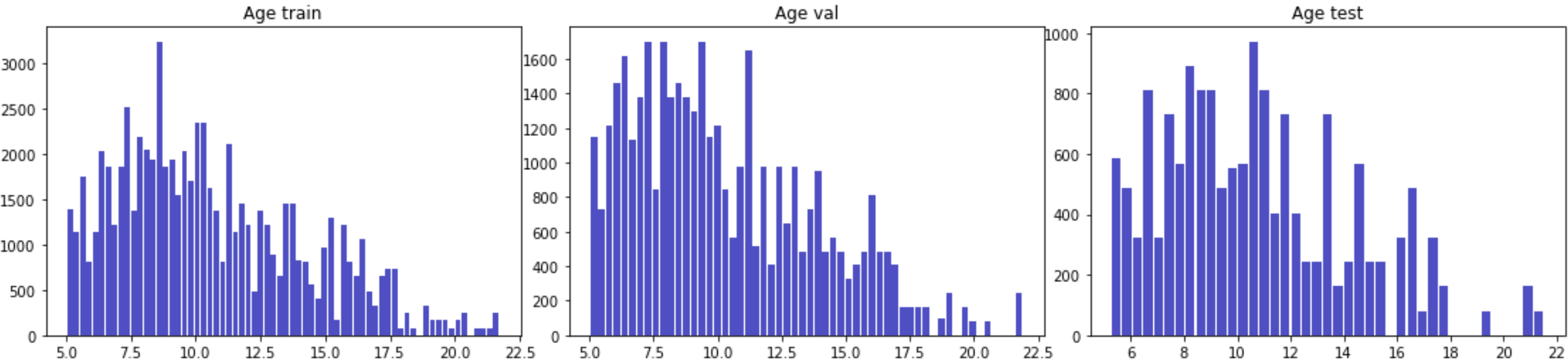


Chromatic image

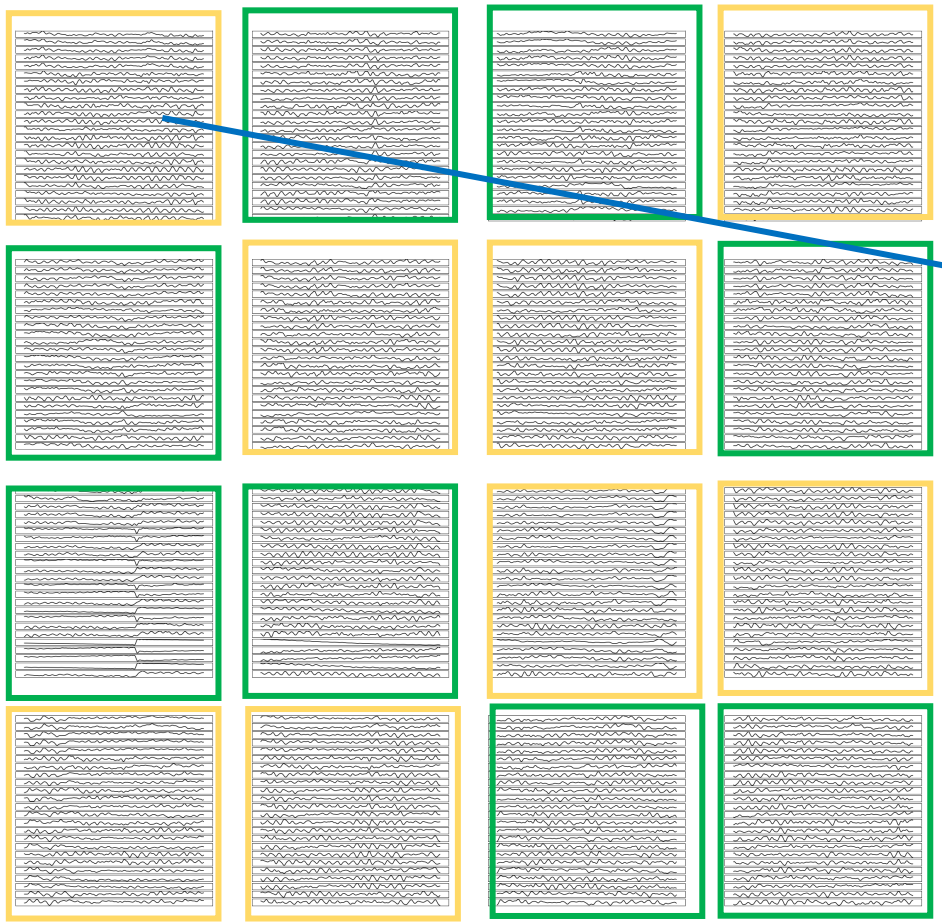


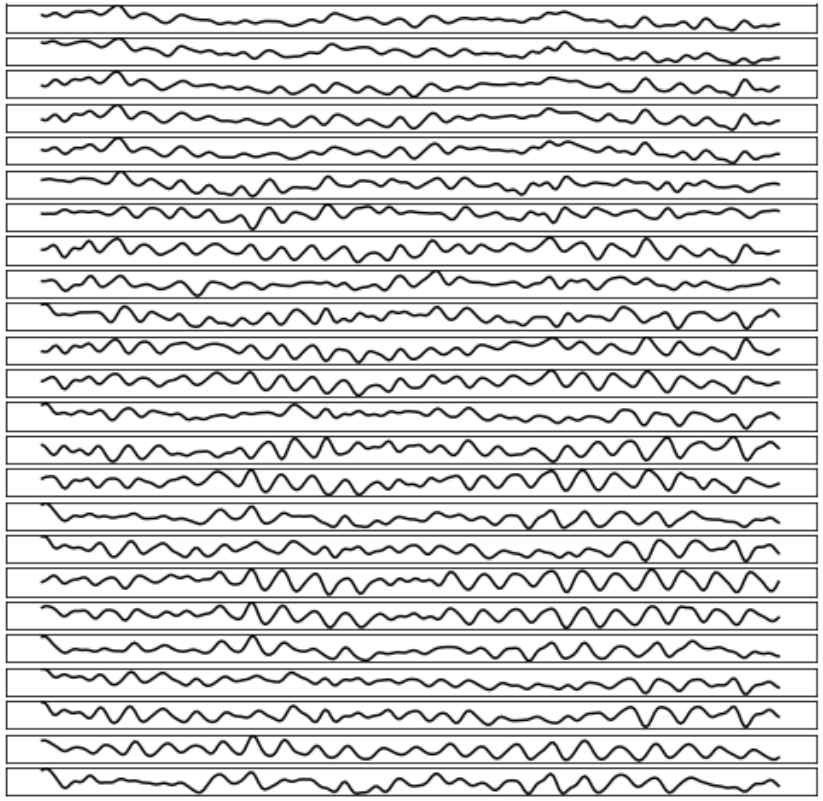
Data

- 10-30-60 split
  - 885 subjects for training -> 71,381 samples
  - 492 subjects for validation -> 39,868 samples
  - 197 subjects for testing -> 15,925 samples

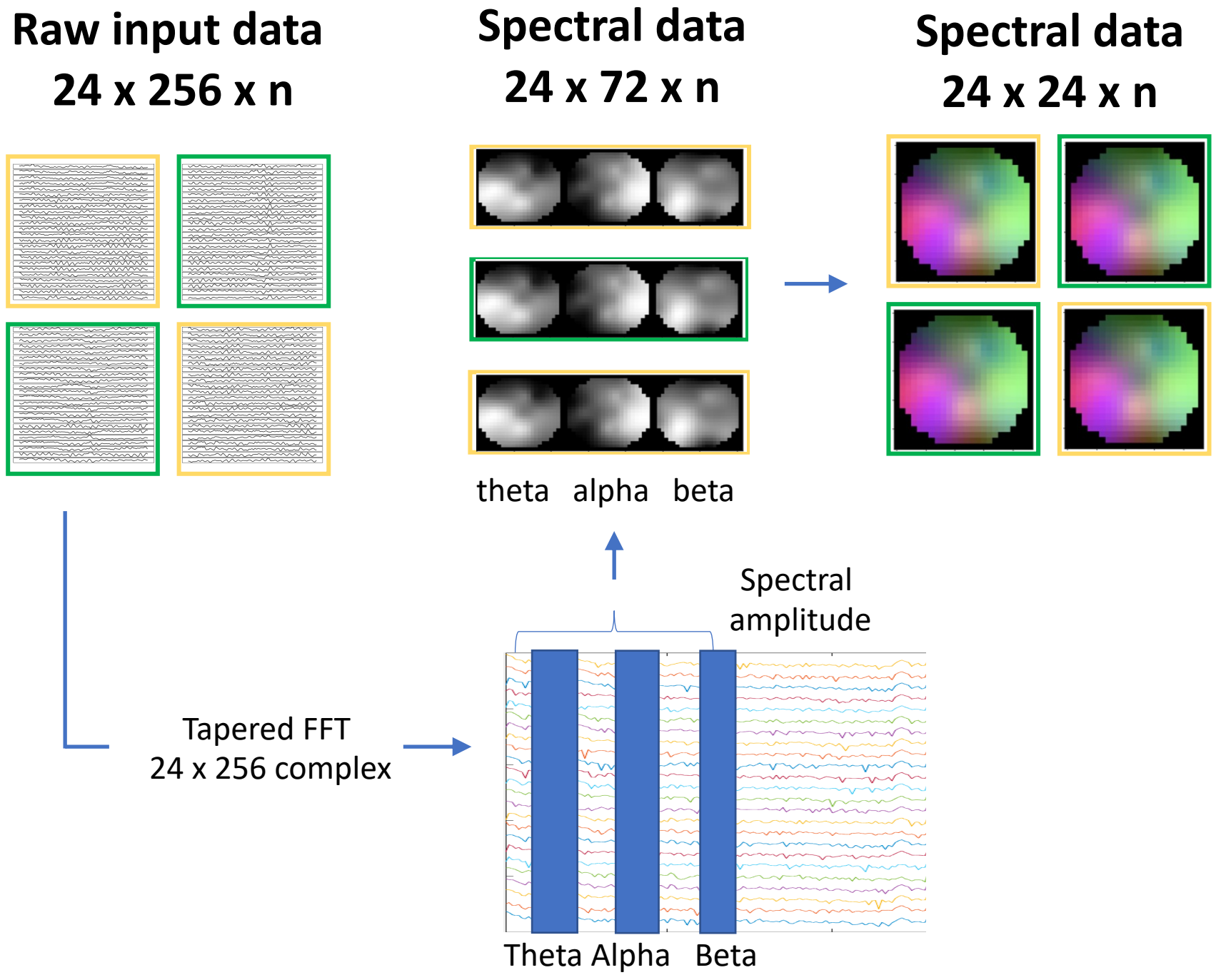


Input data 24 x 256 x n

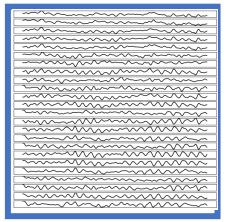




0 for male, 1 for female

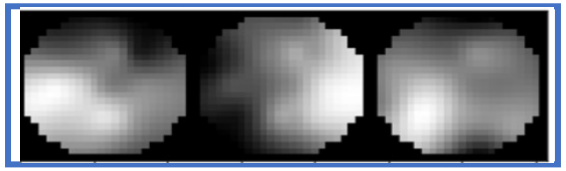

Raw input data
24 x 256 x n
Spectral data
24 x 72 x n
Spectral data
24 x 24 x n
theta alpha beta
Spectral
amplitude
Tapered FFT
24 x 256 complex
Theta Alpha Beta

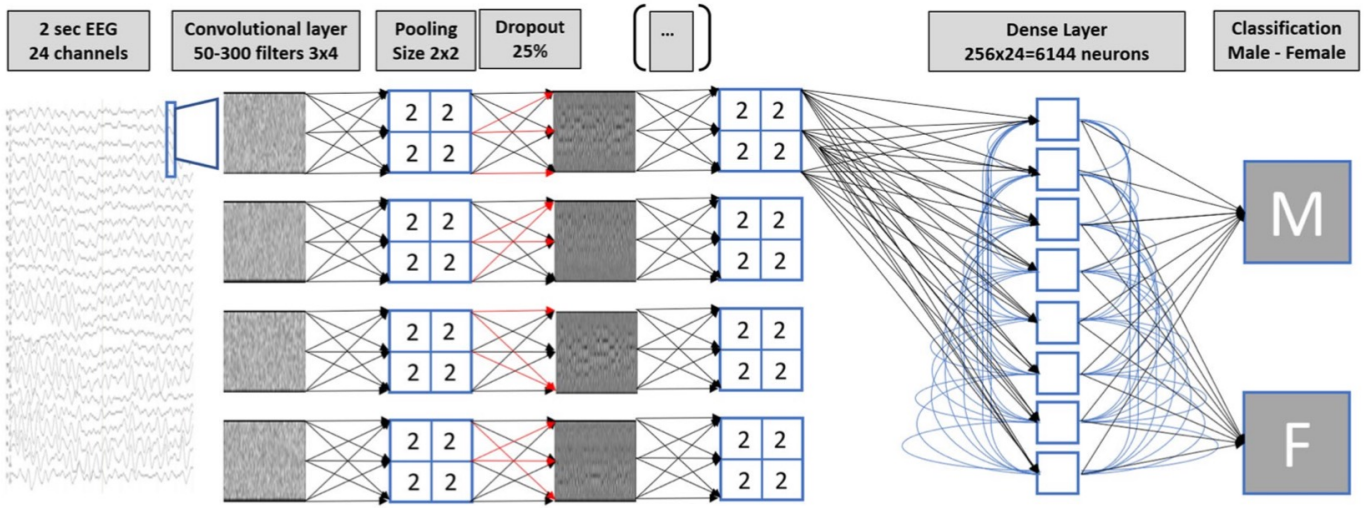
R-SCNN



Or

S-SCNN

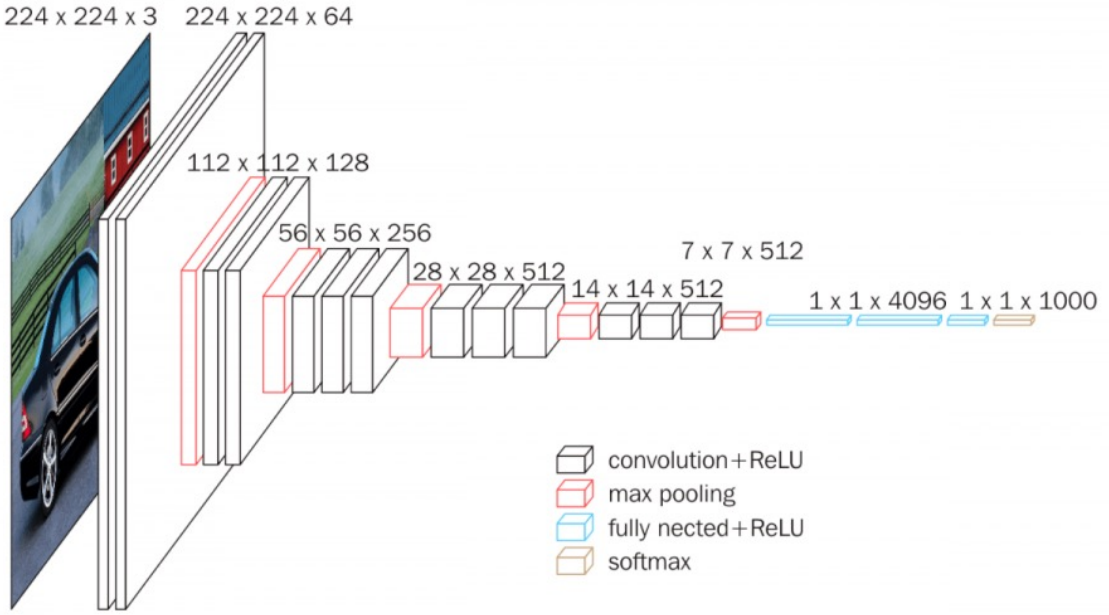




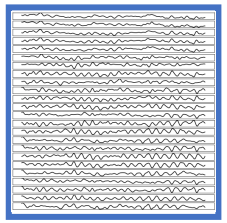
Layer	Filter size	# of filters/hidden units
Convolutional	3x3	100
MaxPooling Dropout (25%)		
Convolutional	3x3	100
MaxPooling Dropout (25%)		
Convolutional	2x3	300
MaxPooling Dropout (25%)		
Convolutional†	1x7	300
MaxPooling* Dropout (25%)		
Convolutional†	1x3	100
Convolutional†	1x3	100
Fully connected		6144
Fully connected		2
Softmax		

Modified VGG16

- Changed input size
- Keep the same scaling between layers
- Number of convolutions divided by 8 for each l
- Dropped layers 19 to 32
- Change number of output classes to 2
- Retrain the whole network

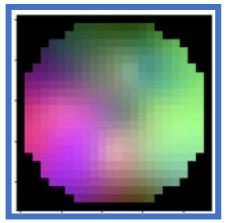


R-VGG



Or

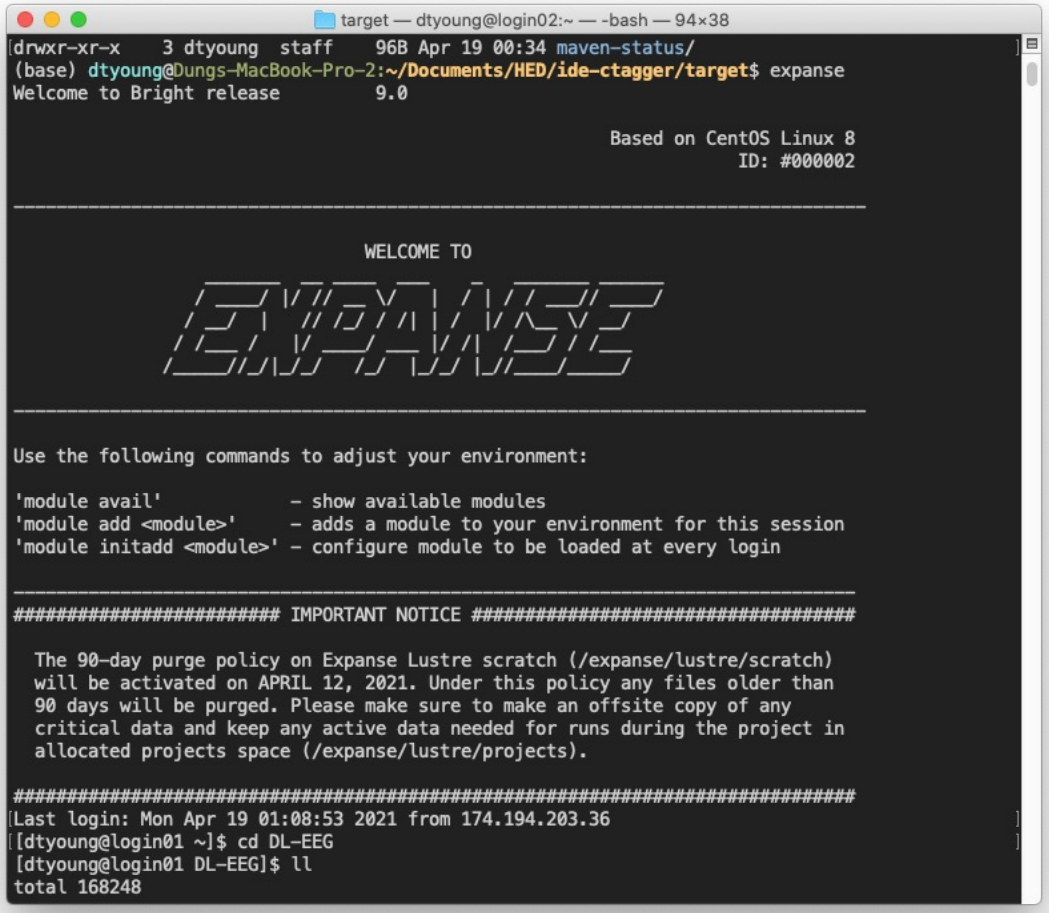
S-VGG



Layer	Filter size	# of filters/hidden units
Convolutional†	3x3	16
Convolutional	3x3	16
MaxPooling		
Convolutional	3x3	32
Convolutional	3x3	32
MaxPooling		
Convolutional	3x3	64
Convolutional	3x3	64
Convolutional	3x3	64
MaxPooling		
Fully connected		1024
Dropout (50%)		
Fully connected		1024
Dropout (50%)		
Fully connected		2
Softmax		

Training the network

- Computing environment



Training the network

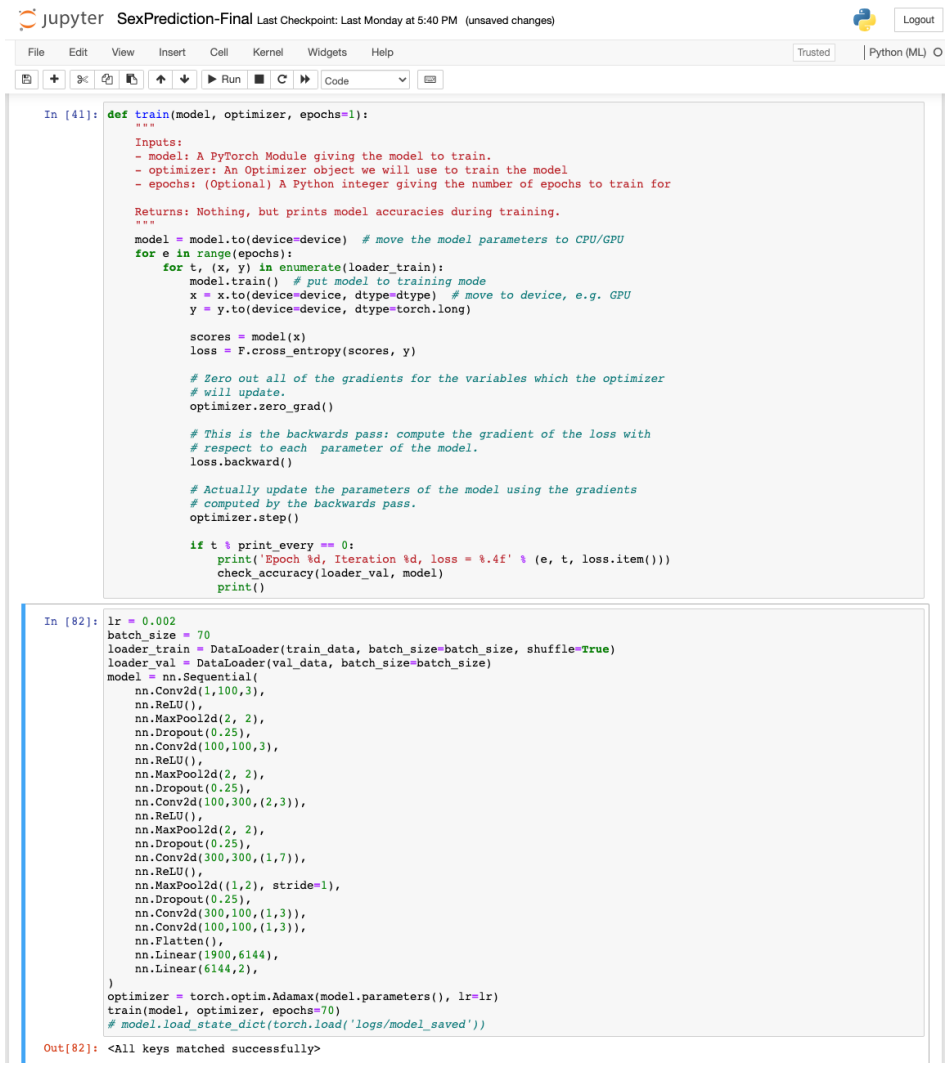
- Programming environment

python

v3.7.10

PyTorch

v1.3.1

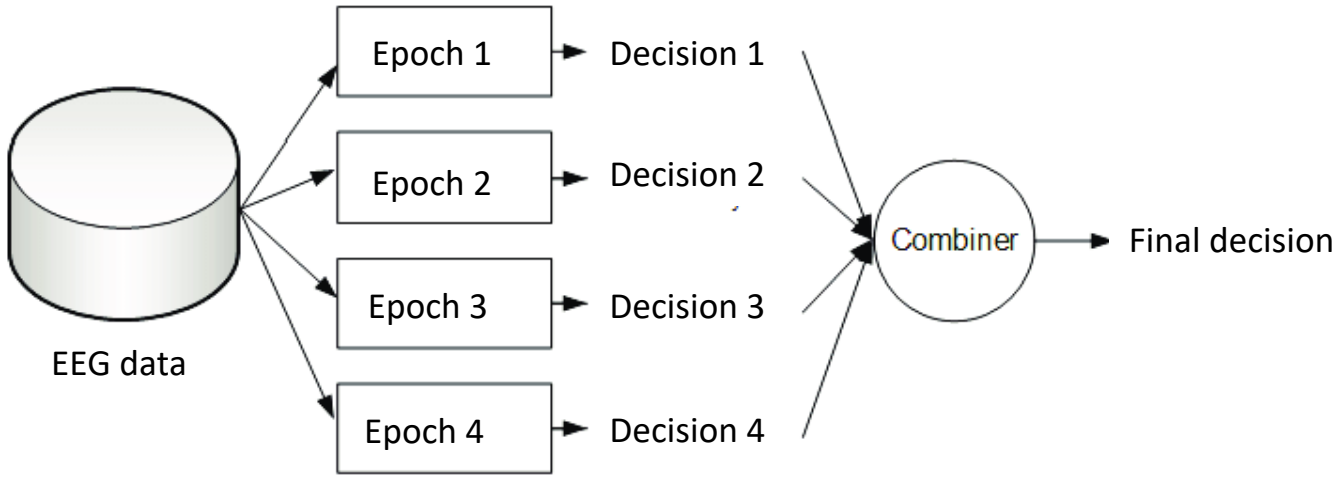


Training the network

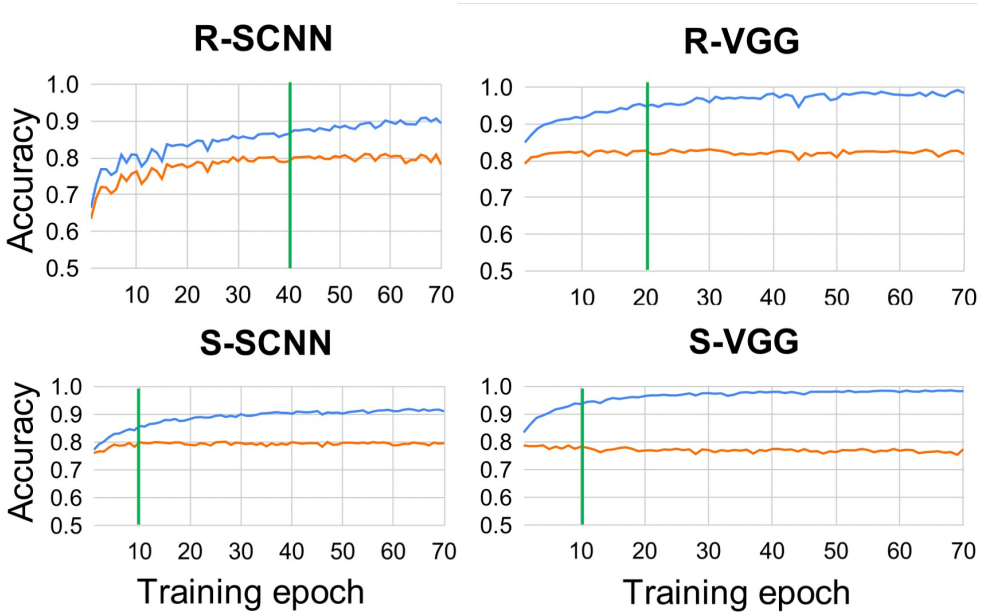
- Adamax optimizer
  - $\beta_1 = 0.9$, $\beta_2 = 0.999$, $\epsilon = 108$ and decay = 0.0
- Learning rate = 0.002
- Batch size = 70 with random shuffle
- 70 epochs

Testing scheme: Majority voting

- Estimates are on 2-second epochs
- We have about 81 samples per subject
- Boost performance by about 4% (see next slide)



Results



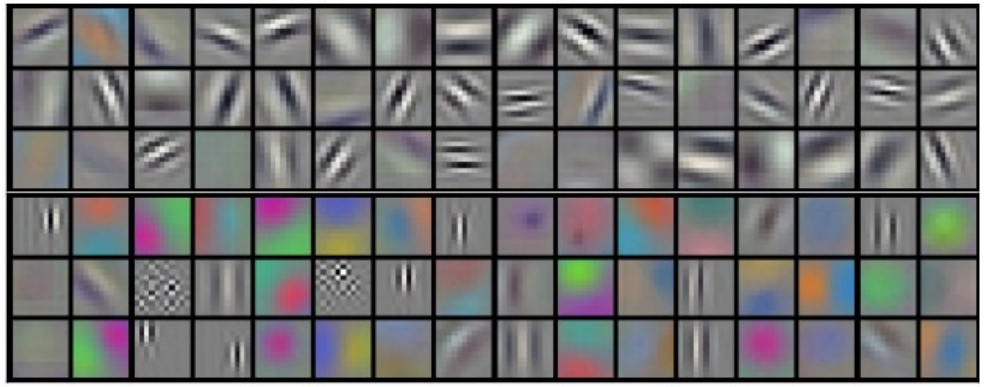
2-sec 2-sec x n with vote

Model	Per-sample	Per-subject
R-SCNN	80.6 (79.7 to 81.5)	85.1 (84.3 to 85.9)
R-VGG	**83.1** (82.7 to 83.4)	**87.0** (86.6 to 87.4)
S-SCNN	79.0 (78.7 to 79.3)	83.2 (82.1 to 84.3)
S-VGG	77.1 (76.8 to 77.4)	81.3 (80.0 to 82.6)

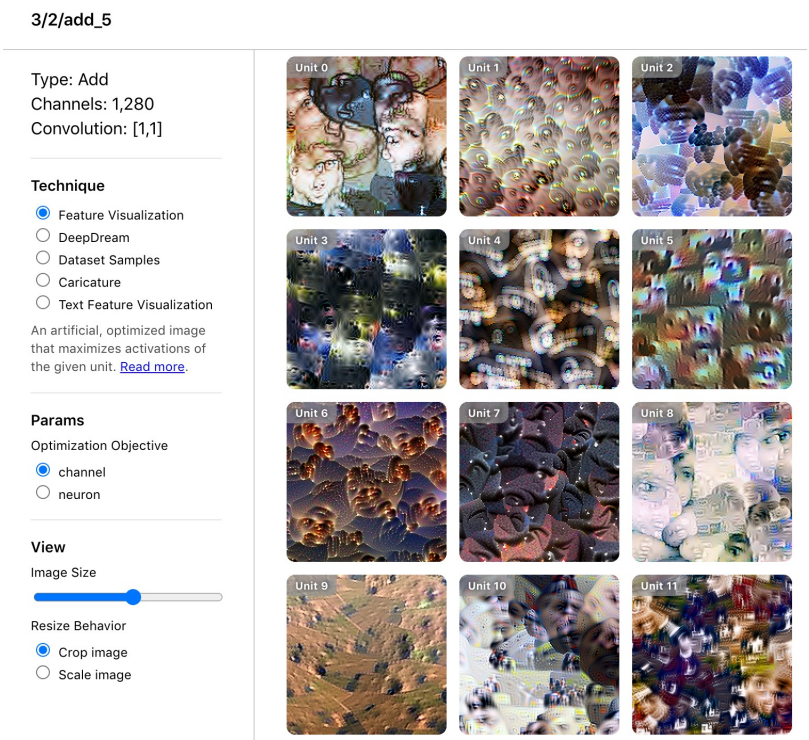
Table 3. Models classification accuracy. 95% confidence interval is indicated in parenthesis. Bolded values indicate best performance.

Truong, D., Milham, M., Makeig, S., and Delorme, A. Deep Convolutional Neural Network Applied to Electroencephalography: Raw Data vs Spectral Feature. Annu Int Conf IEEE Eng Med Biol Soc, 2021, pp. 1039-1042, doi: 10.1109/EMBC46164.2021.9630708.

Assessing features learned by the network



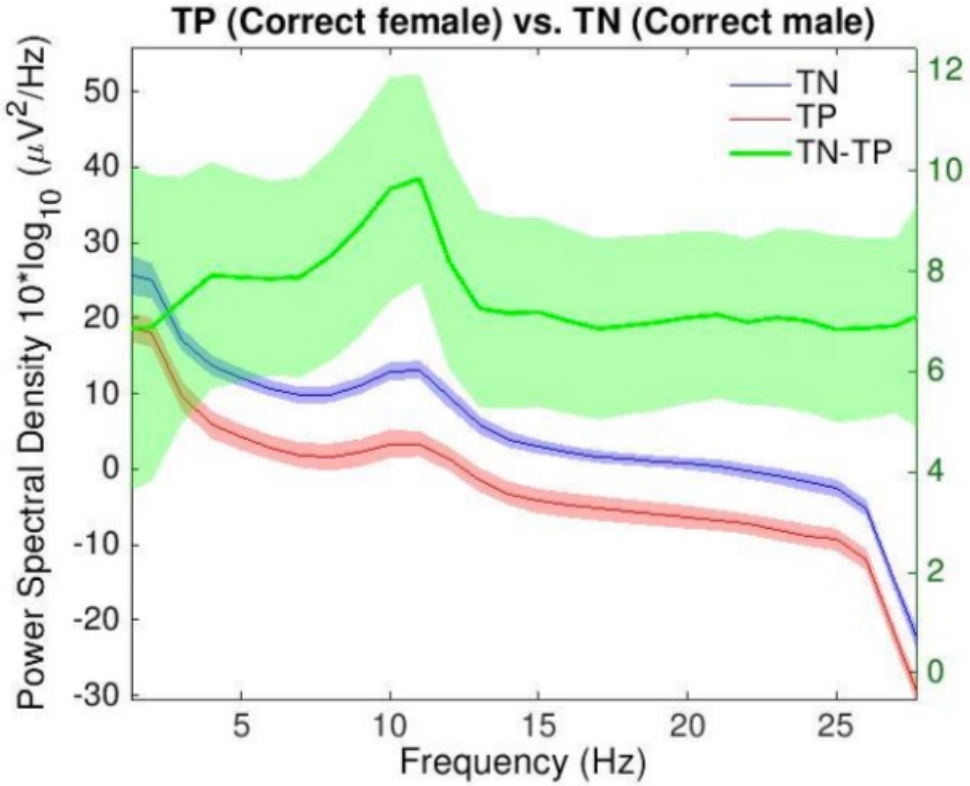
Krizhevsky et al., 2012



Best samples

- "Best" = gives highest activation in classification neurons
- Get samples of the top 20 subjects in validation set
- Best male samples show higher spectral power across frequencies, most notably the alpha band near 10Hz

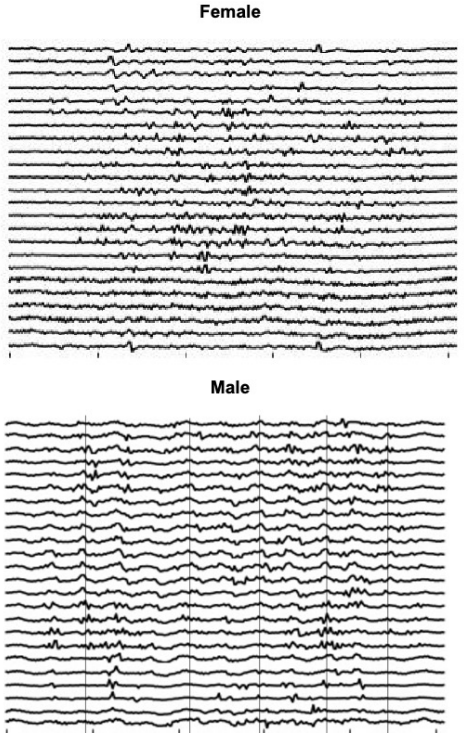
	Classified as female	Classified as male
Female sample	**True Positive (TP)**	False Negative (FN)
Male sample	False Positive (FP)	**True Negative (TN)**

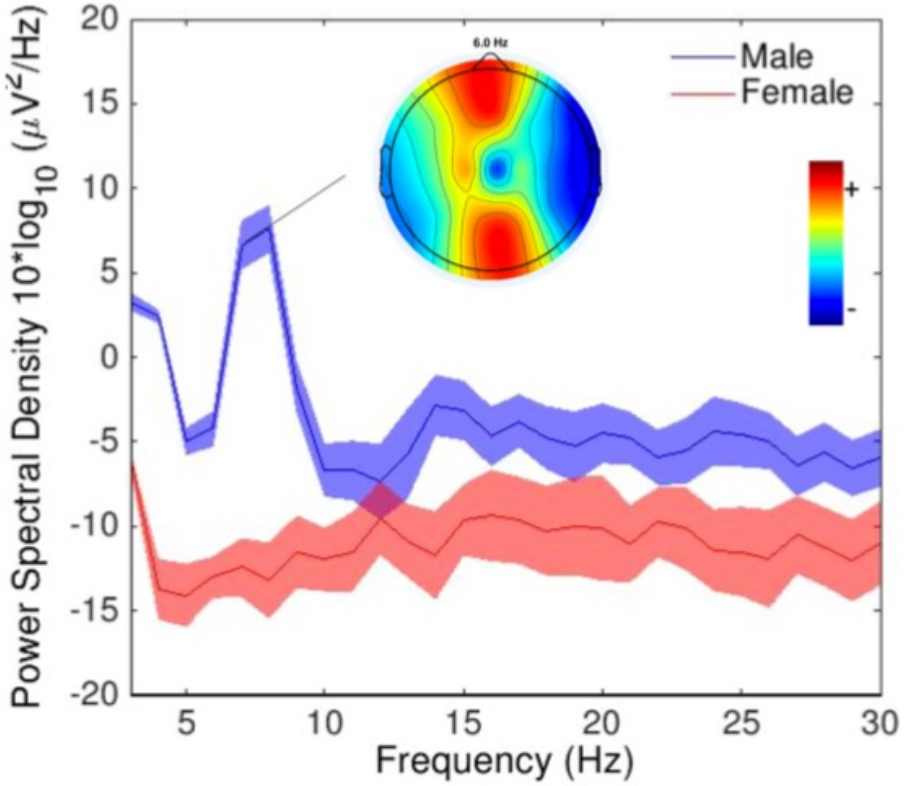


Truong, D., Makeig, S., & Delorme, A. (2021). Assessing learned features of Deep Learning applied to EEG. *arXiv preprint arXiv:2111.04309*.

Activation Maximization

- Synthesize the input that maximize the activation of the classification neurons
- Get 20 samples for each sex
- Best male samples show distinctly higher high theta power for male samples (6-8 Hz)

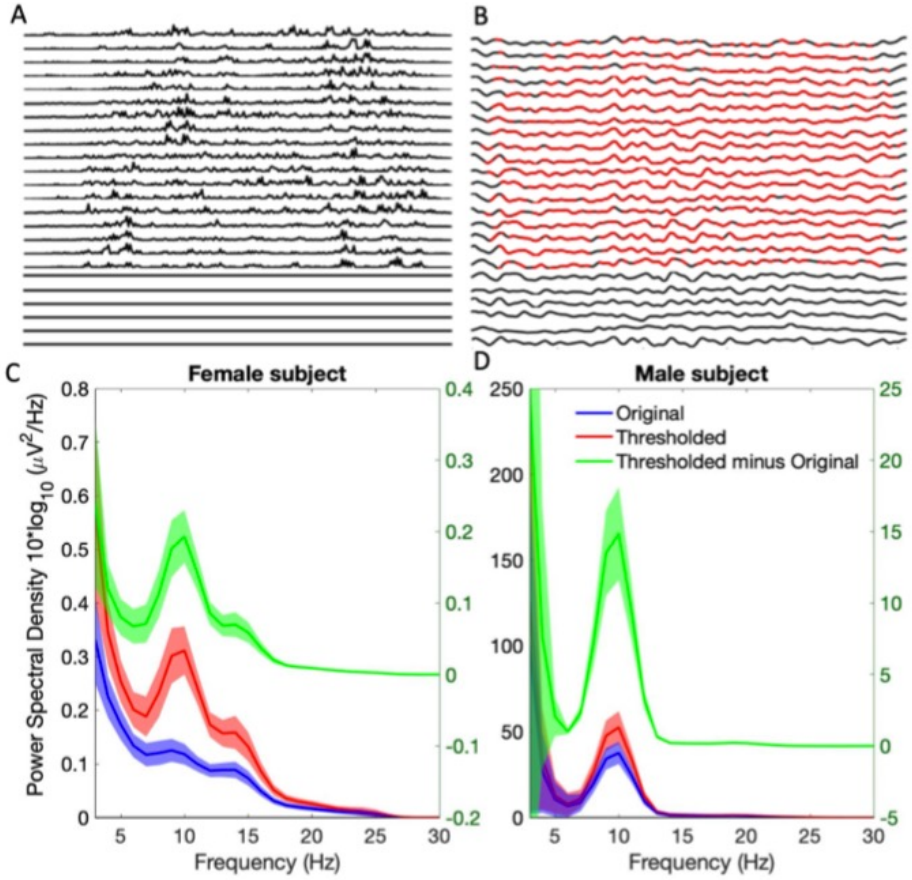




Truong, D., Makeig, S., & Delorme, A. (2021). Assessing learned features of Deep Learning applied to EEG. *arXiv preprint arXiv:2111.04309*.

Saliency map

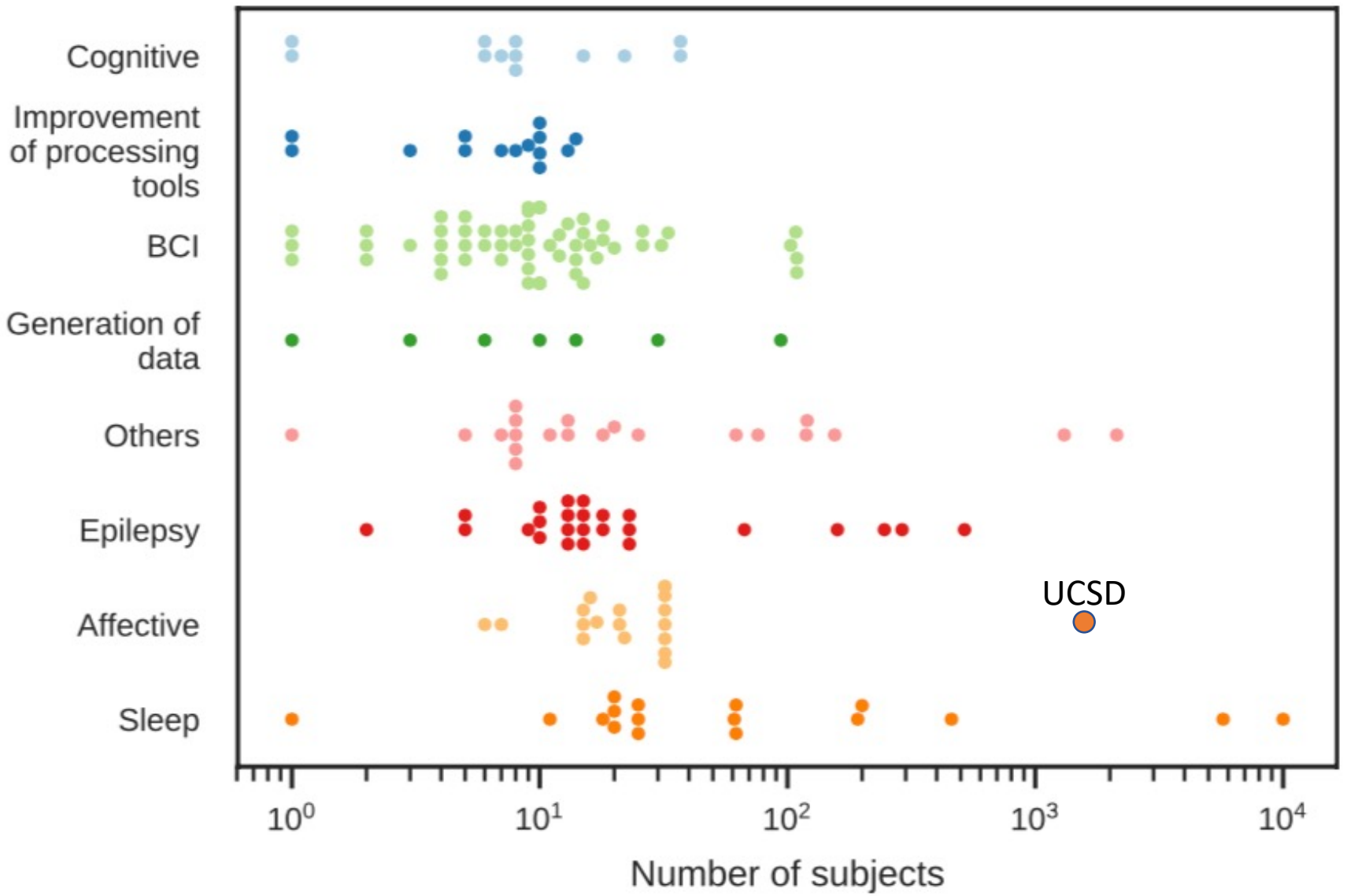
- Back-project the gradient of the classification neuron to the input
- Magnitude of the gradients for each input value indicate the importance of that value to the neuron's activation
- Raw EEG samples contributing the least to the classification (gradients fell below the 30% quantile threshold) were removed then linearly interpolated using the remaining samples
- Thresholded samples showed higher alpha power, most notably near 10Hz



Truong, D., Makeig, S., & Delorme, A. (2021). Assessing learned features of Deep Learning applied to EEG. *arXiv preprint arXiv:2111.04309*.

Discussion

- Most EEG datasets are order of magnitude smaller



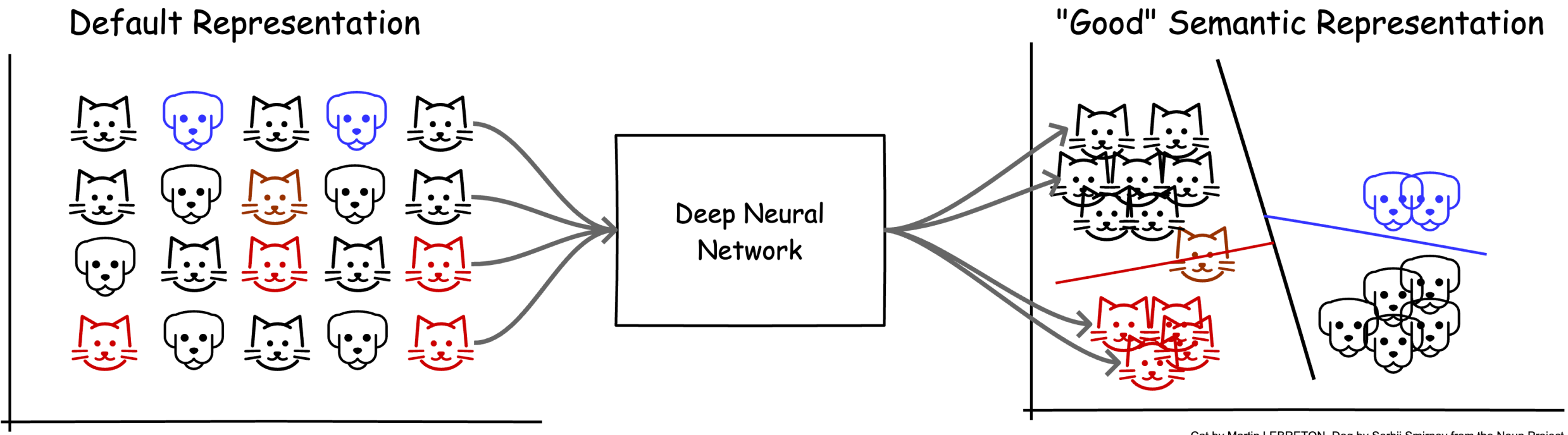
Roy Y. et al (2019)

Discussion

- Most EEG datasets are order of magnitude smaller
- Not generalizable to different number of channels and sampling rate
- Problem chosen could be simple: boys move more than girls, known sex difference in skull thickness
- VGG-16 is no longer state-of-the-art model nor designed specifically for time-series/EEG data
- Scientific values of learned features visualization remain unclear

Future work: representation learning

- DL models take data from the original space and map it to a “meaningful” representation

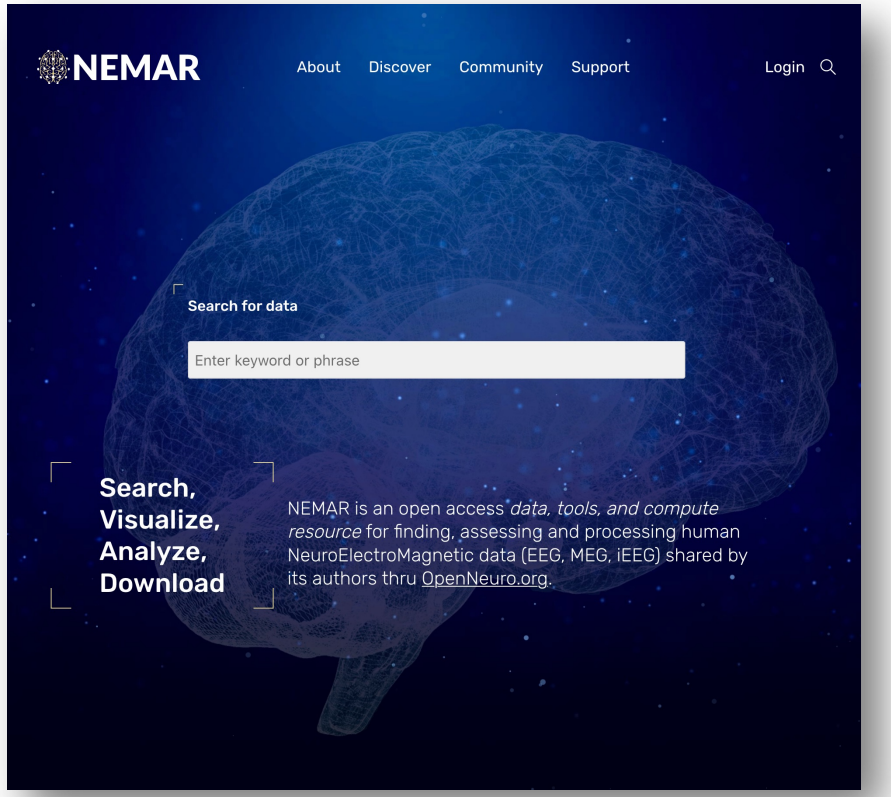


Cat by Martin LEBRETON, Dog by Serhii Smirnov from the Noun Project

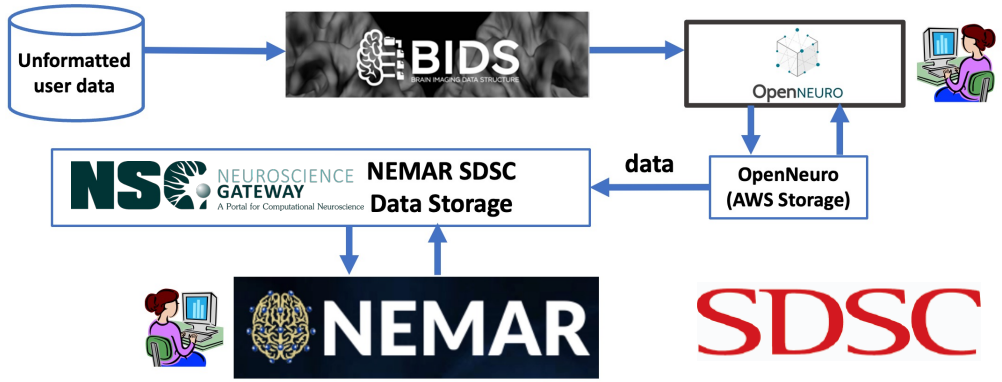
Future work: representation learning

- DL models take data from the original space and map it to a “meaningful” representation
- Can we learn embeddings that are useful across tasks?
- Can we leverage embeddings across modalities to study naturalistic cognition?
- What network architectures and training regimes can generalize across datasets?

Data, tools, and compute resources (DATCOR)



https://nemar.org

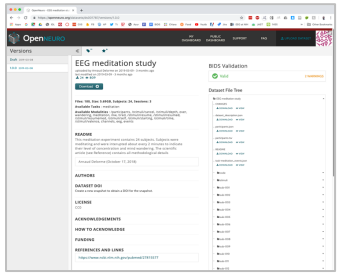


Total dataset number	128 (5TB)
Number of EEG dataset	95
Number of MEG dataset	20
Number of iEEG dataset	13
Hours of recording	3,000+
Number of participants	6,827

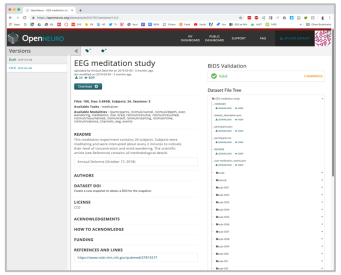
Task names from NEMAR datasets

Standardized event description enables analysis across datasets

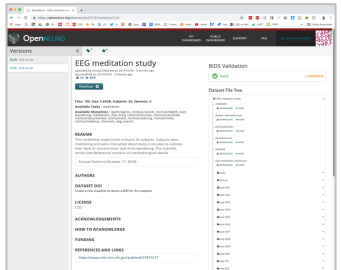
OpenNeuro/NEMAR BIDS #1



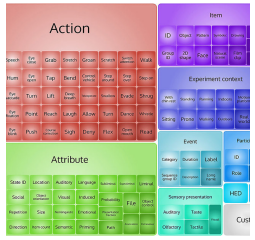
OpenNeuro/NEMAR BIDS #2



OpenNeuro/NEMAR BIDS #3

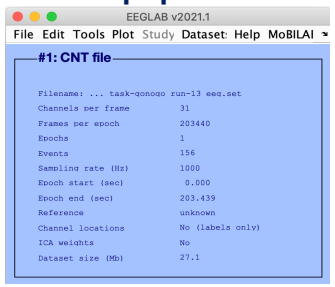


HED event selection

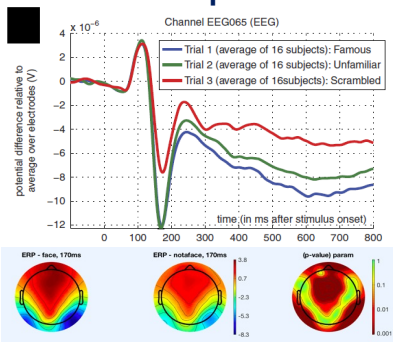


https://hedtags.org

EEGLAB automated pipeline



Output



Future development: *BIDS-DL* plug-in

- Automated pipeline to convert HED-identified data segments from BIDS datasets to DL-ready dataset
- Host data on the cloud and make it streamable so that no data download/upload required

Truong, D., Sinha, M., Venkataraju, K. U., Milham, M., & Delorme, A. (2022) A streamable large-scale clinical EEG dataset for Deep Learning. The 44th International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC); July 11-15, 2022.

Thank you