Neural networks and deep learning

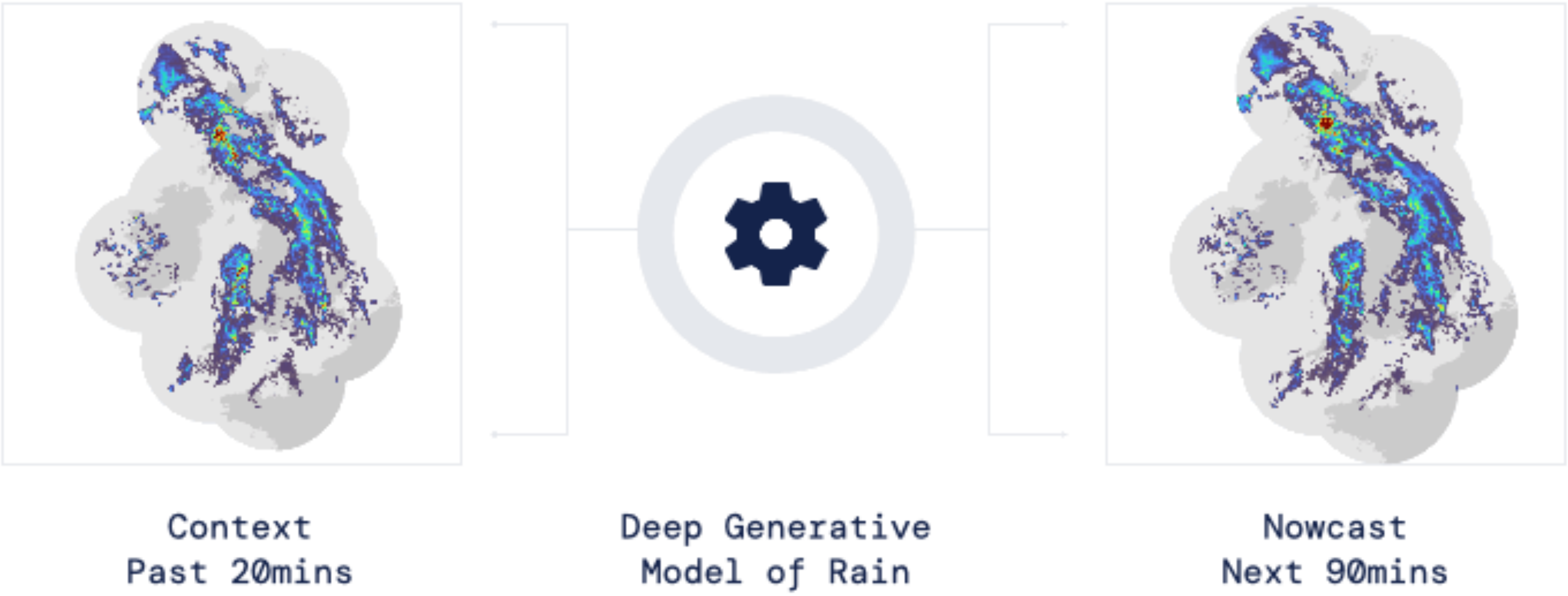
Deep Machine Learning in Weather and Climate

Jesper Dramsch

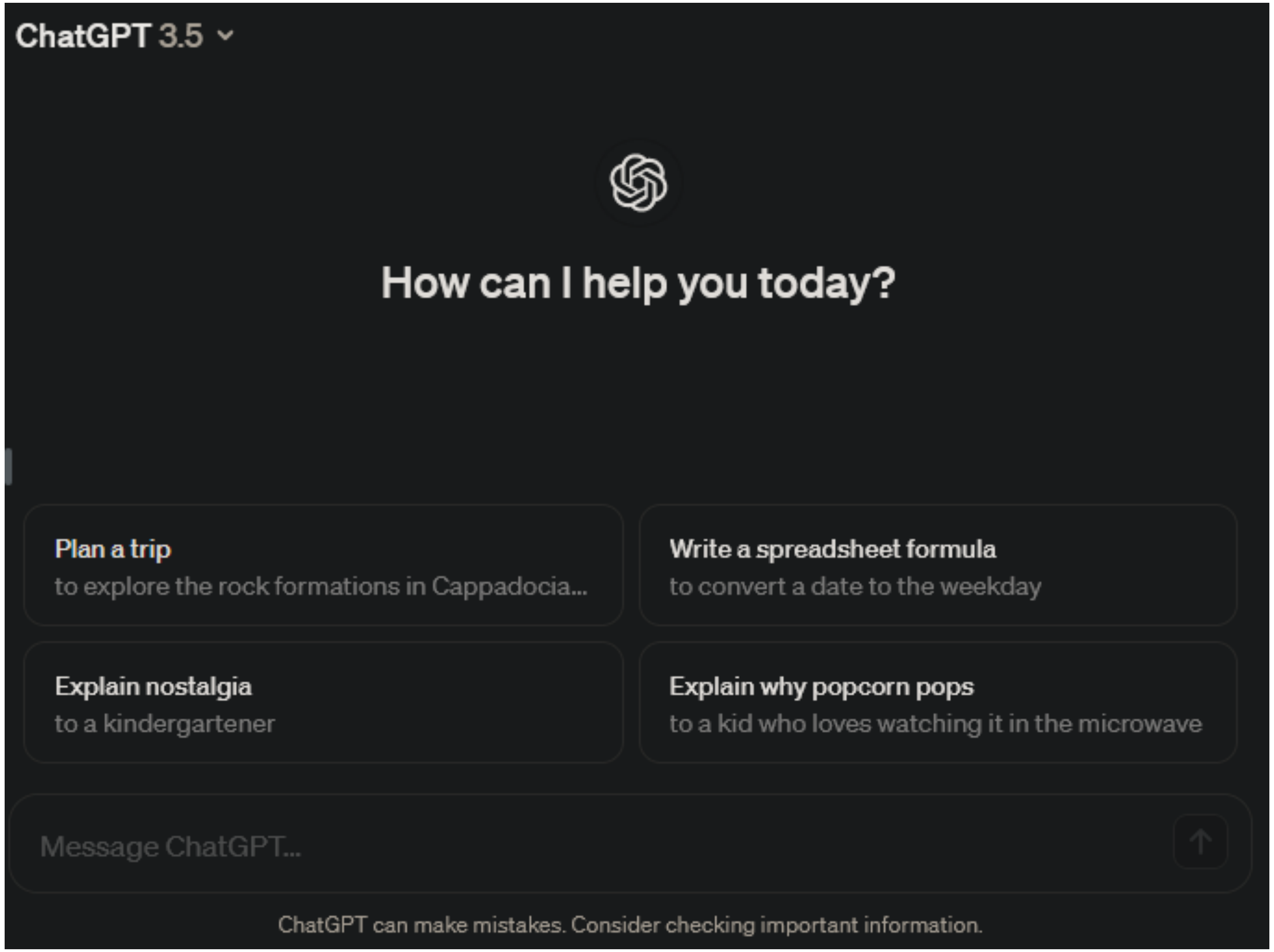
ECMWF Bonn

Jesper.Dramsch@ecmwf.int

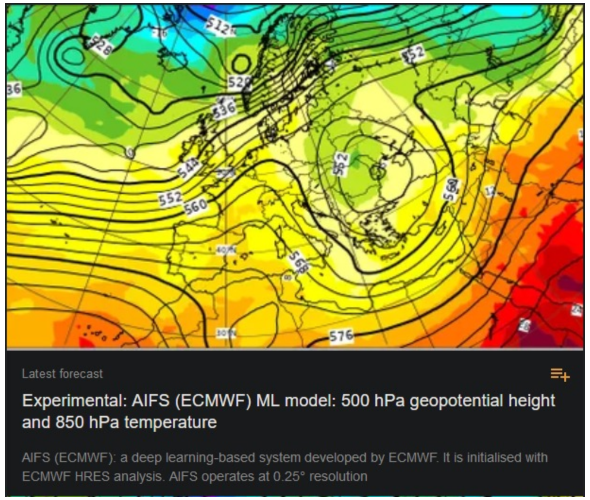
Deepmind Nowcasting predicting the future

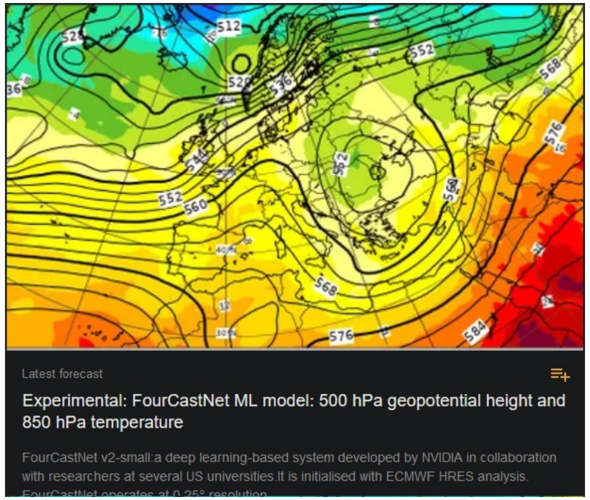


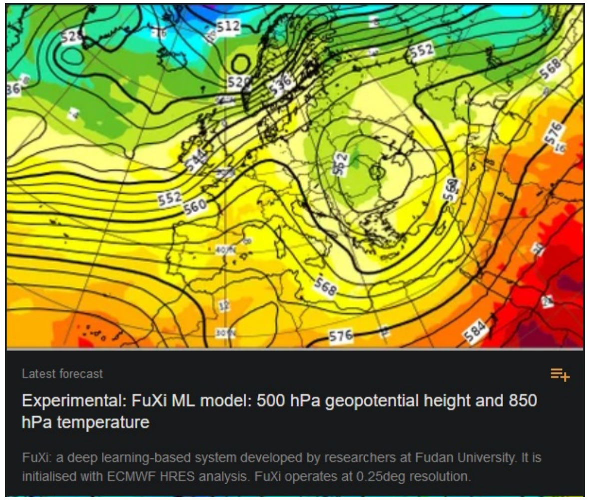
ChatGPT happened...

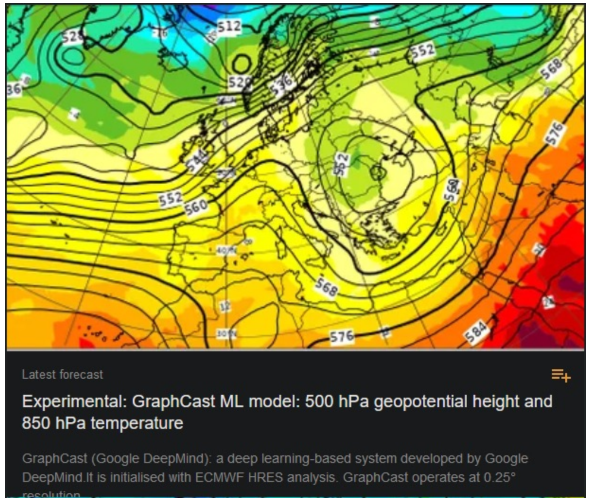


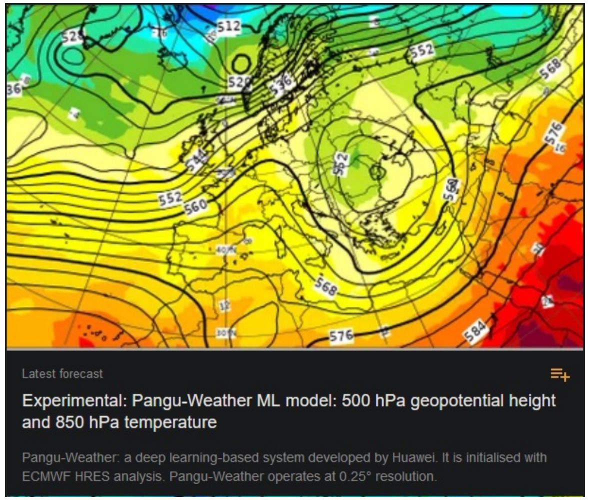
The Rise of Data-Driven Weather Forecasting









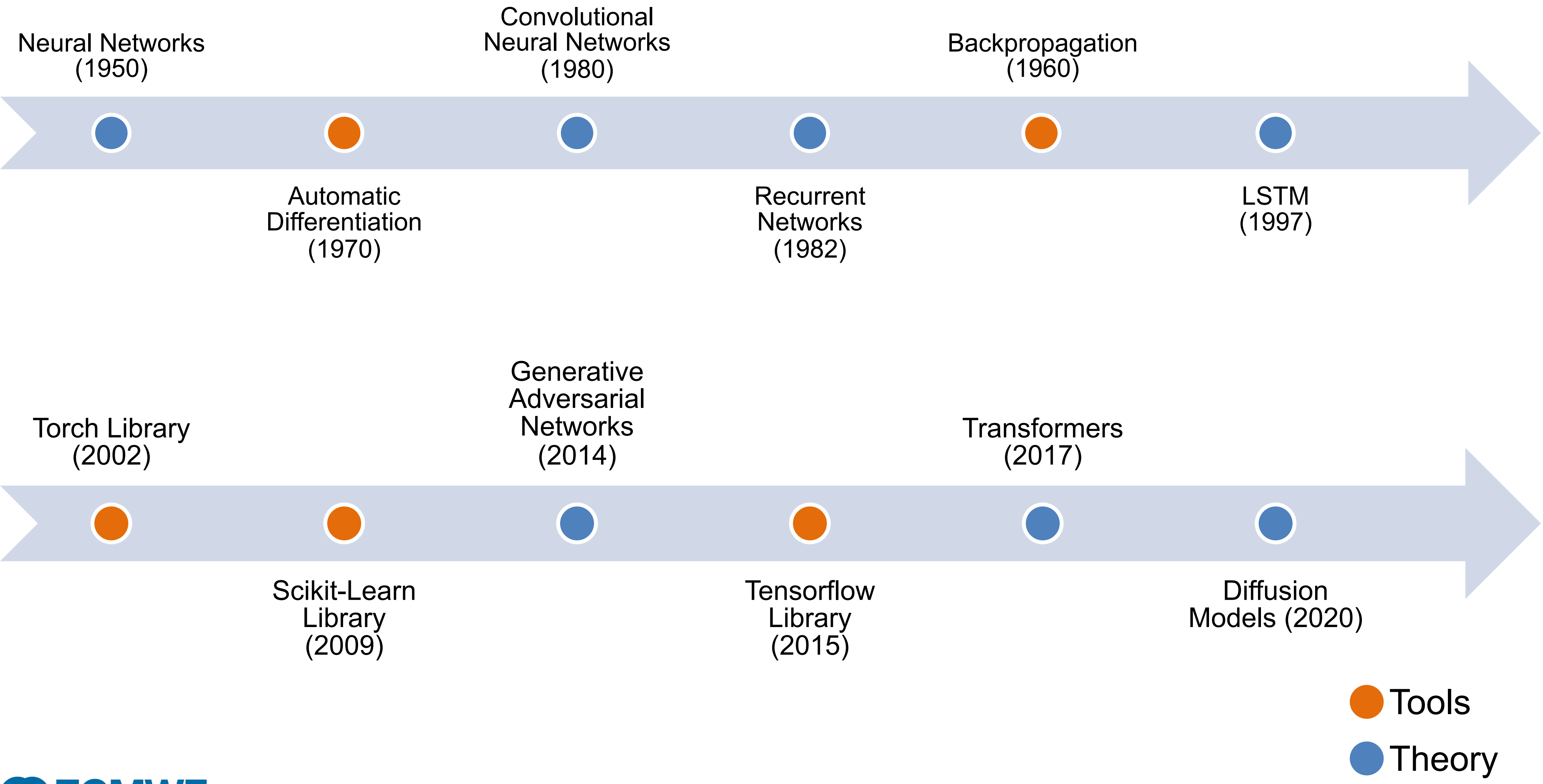


Outline

- Quick history of neural networks
- Dense networks
- Neural network training and GPUs
- Convolutional neural networks
- Recurrent networks
- Transformers

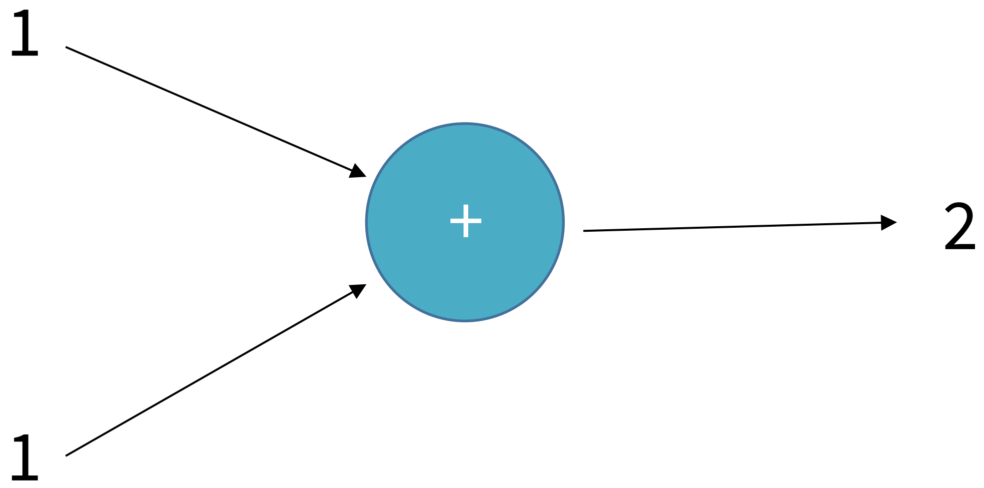
Quick history of neural networks

A Short History of Neural Networks



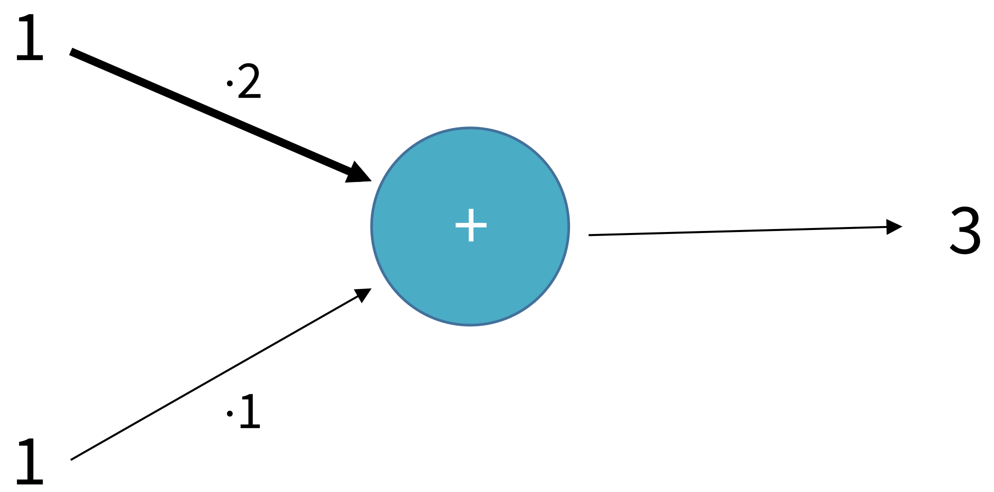
Dense networks

A Simple Neuron for Addition

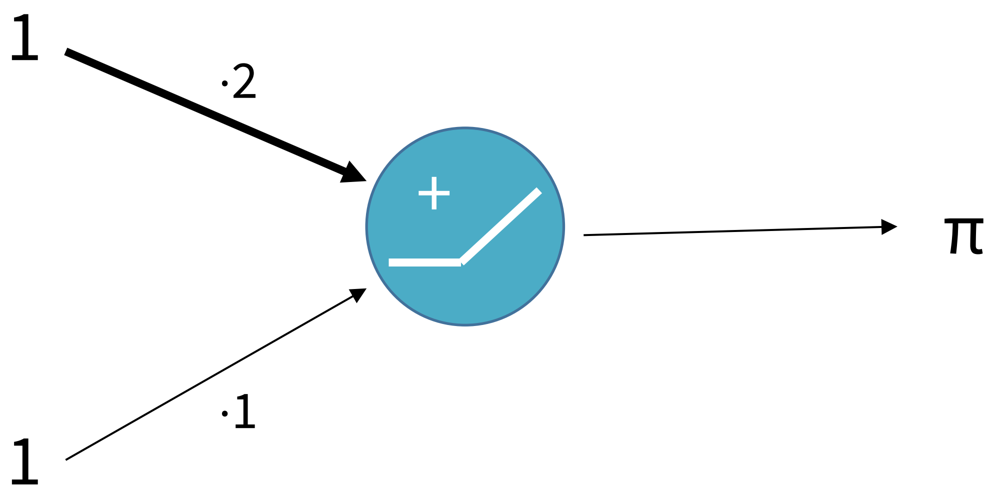


ECMWF

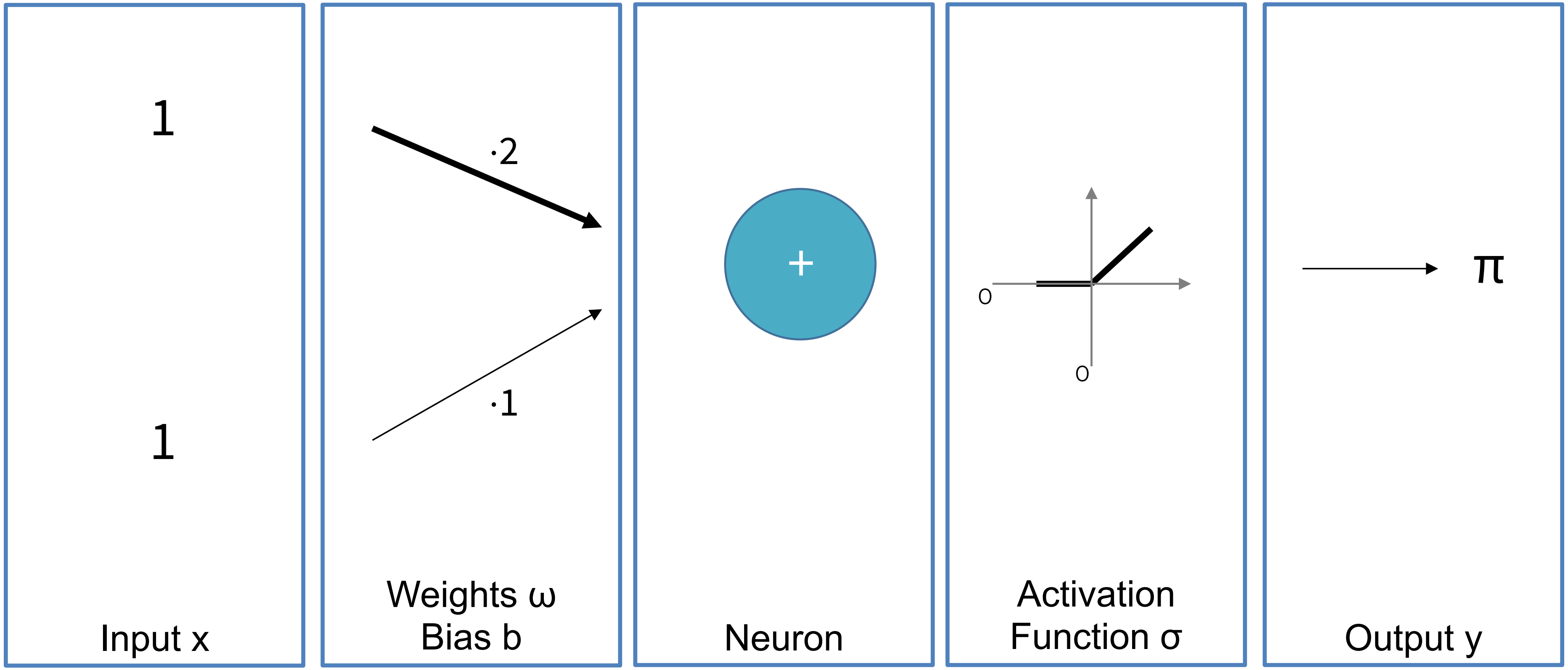
A Simple Neuron – Changing Weights



A Simple Neuron – Activation Function



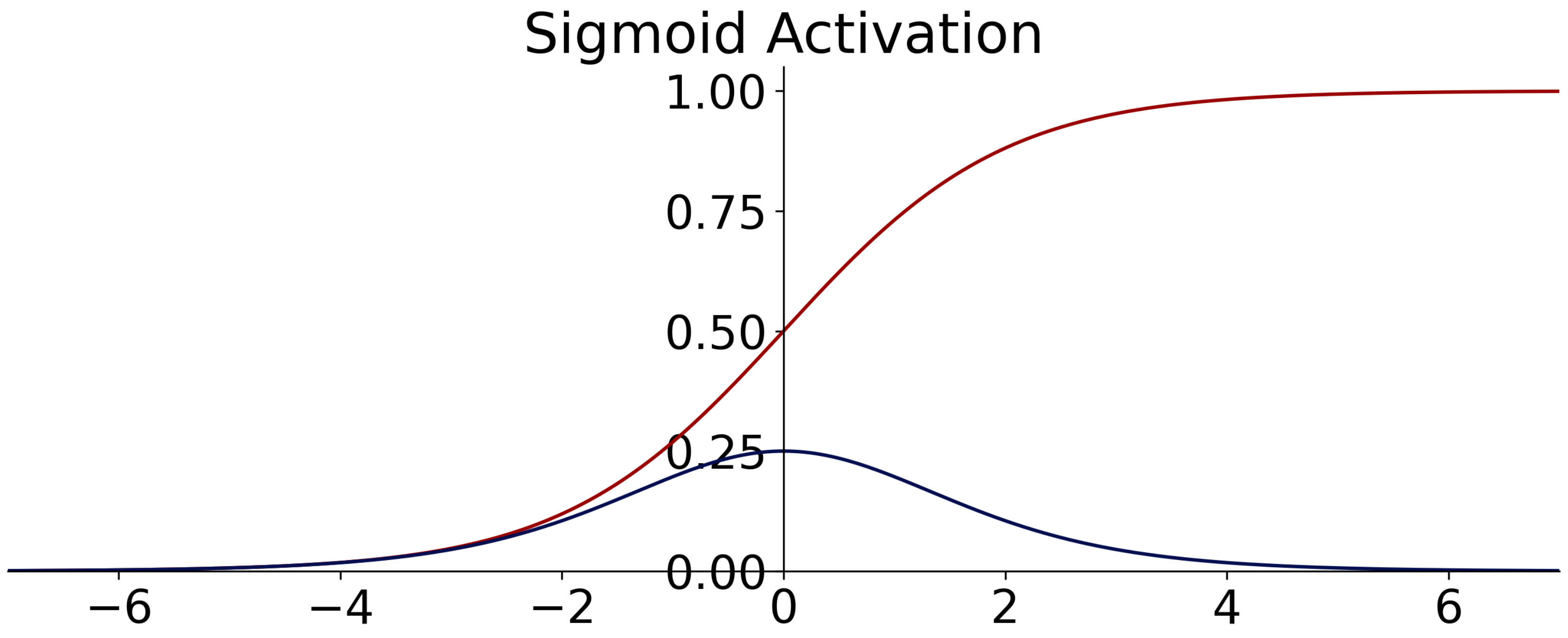
A Simple Neuron – Deconstructed



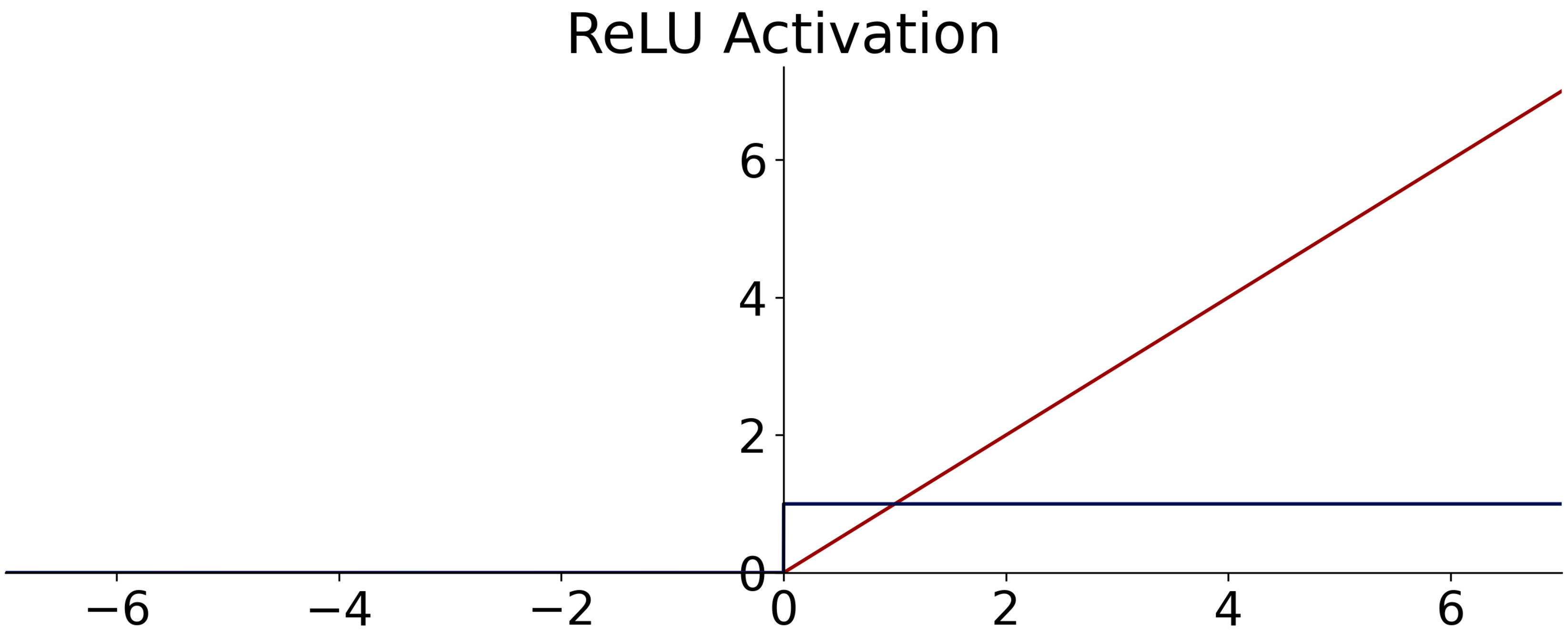
$$y = \sigma\,(\,\omega \cdot x + b)$$

ECMWF

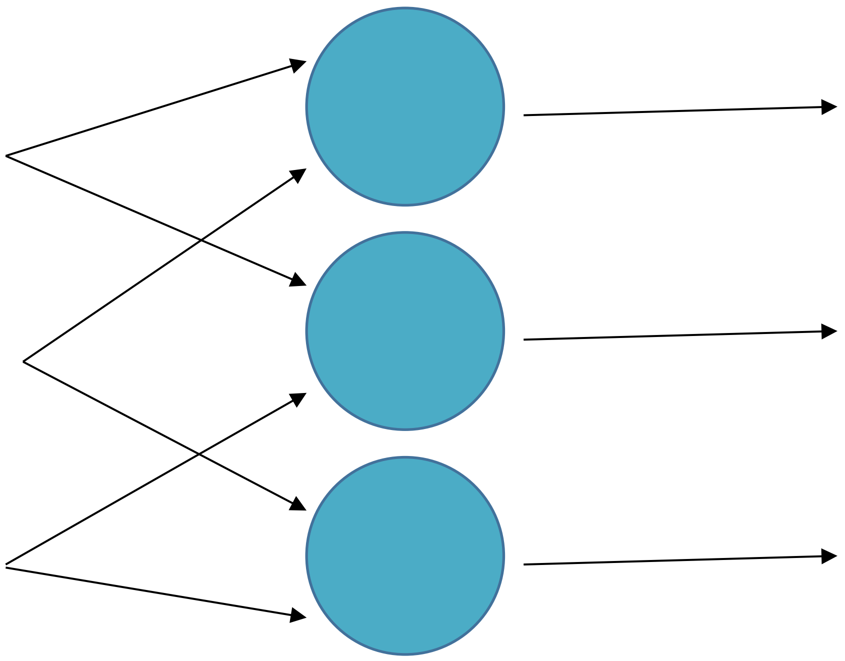
Classic Activation Function



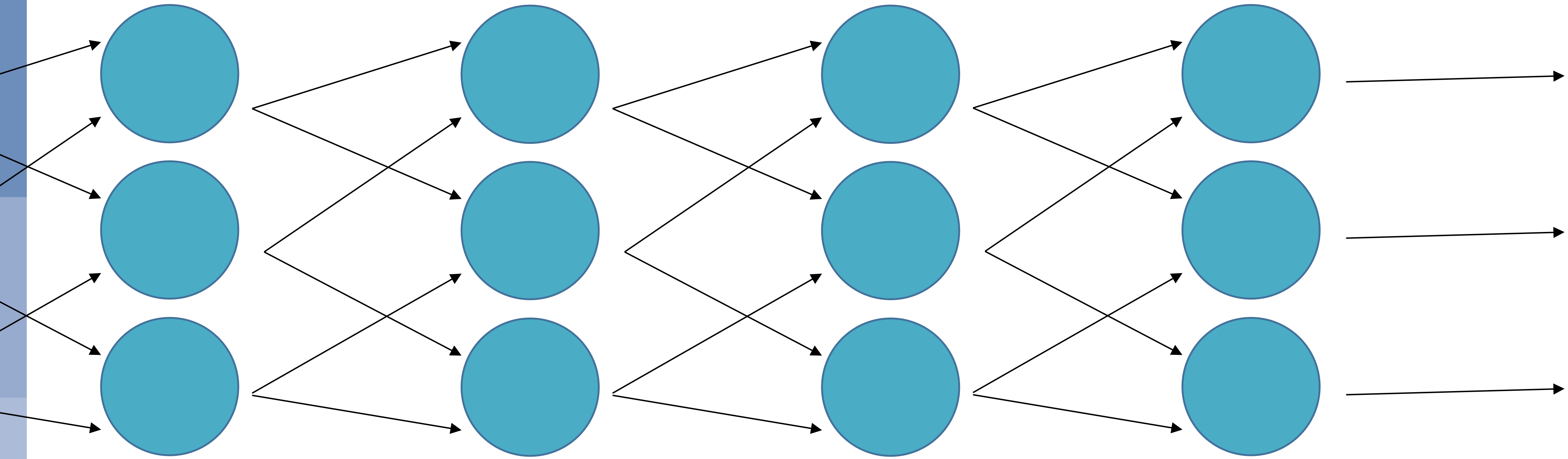
Modern Activation Functions



A Small Neural Network

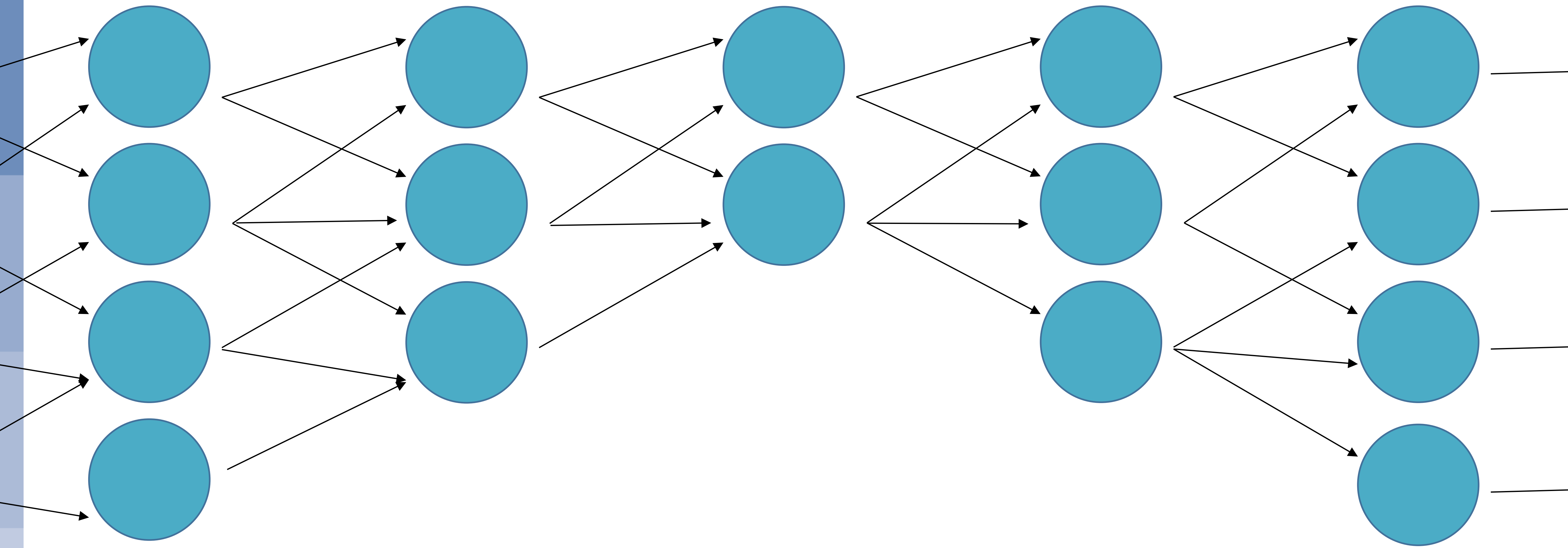


A Deep Neural Network



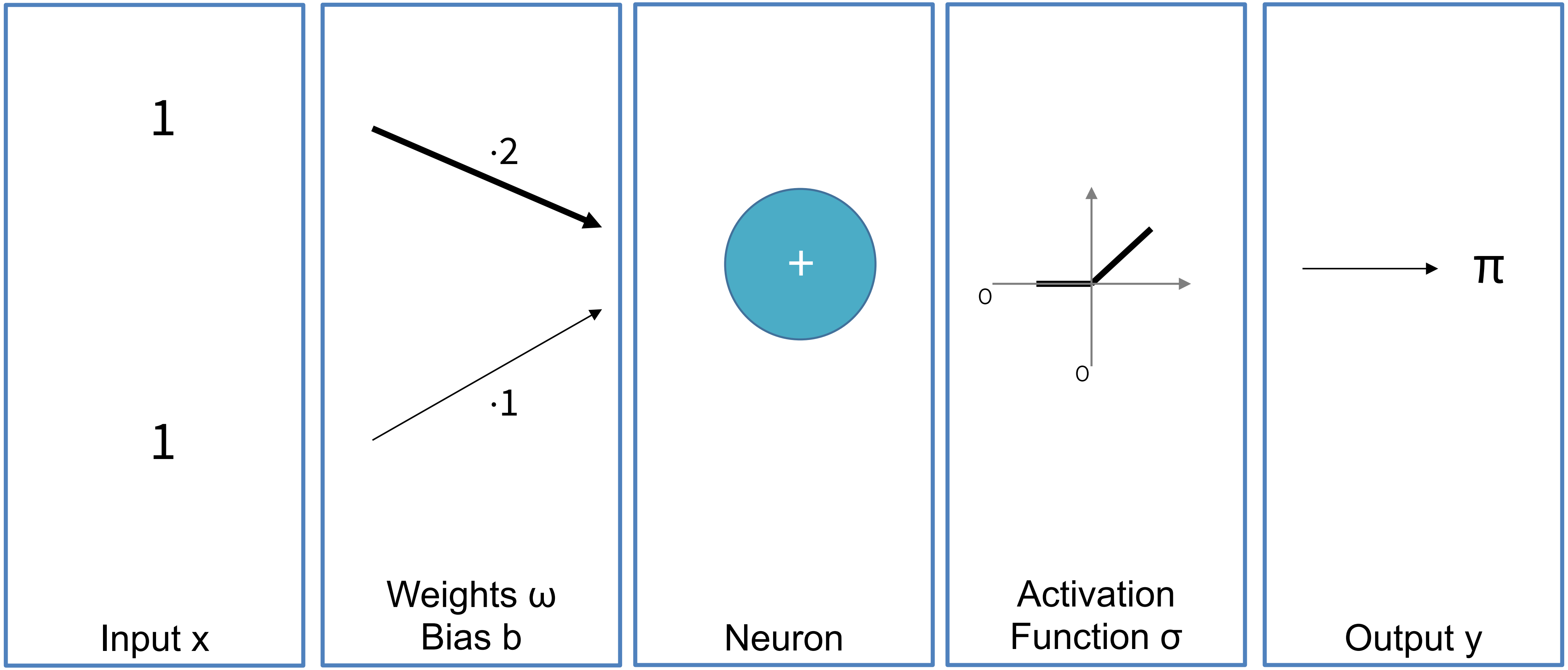
ECMWF

Different Combinations In Neural Networks



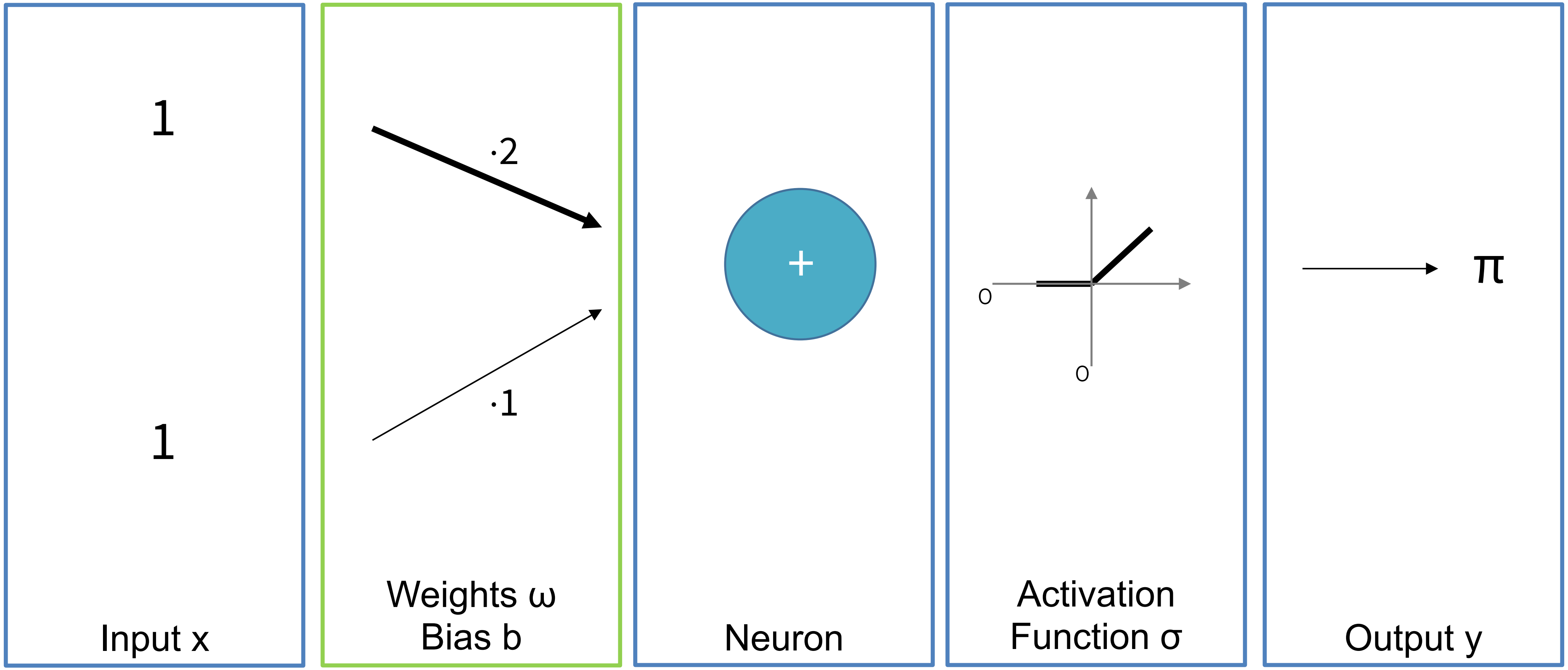
Neural network training

Learnable Parameters



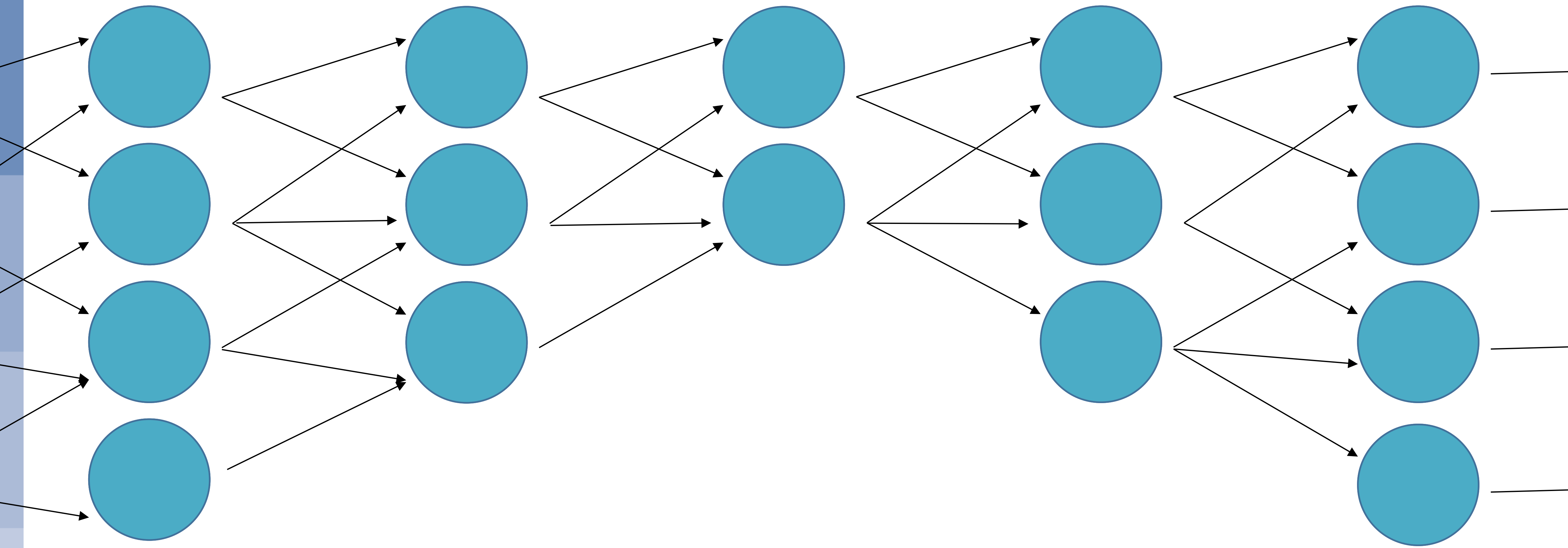
$$y = \sigma(\omega \cdot x + b)$$

Learnable Parameters



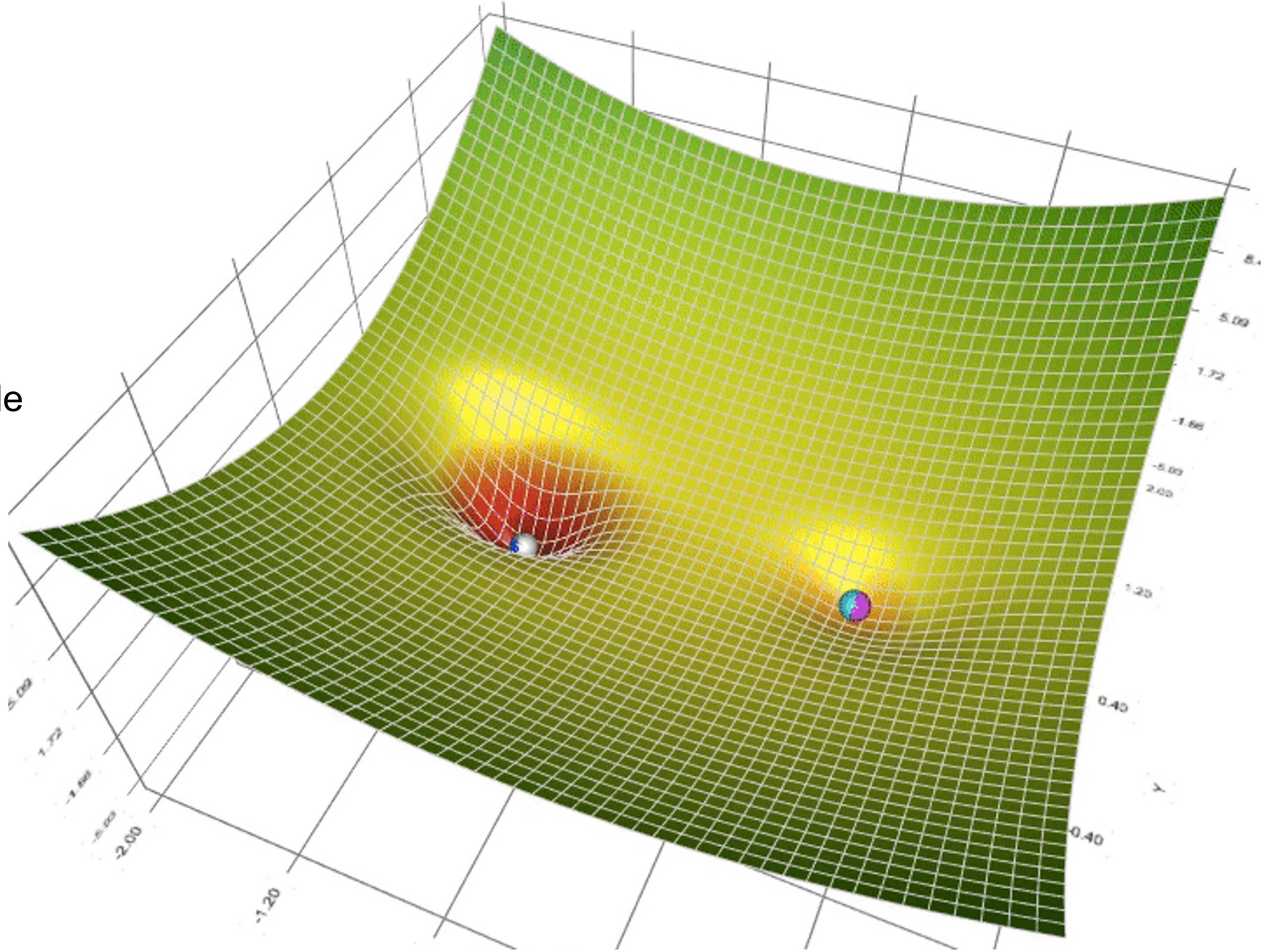
$$y = \sigma\,(\,\boldsymbol{\omega} \bullet x + \mathbf{b})$$

Forward Pass In Neural Networks



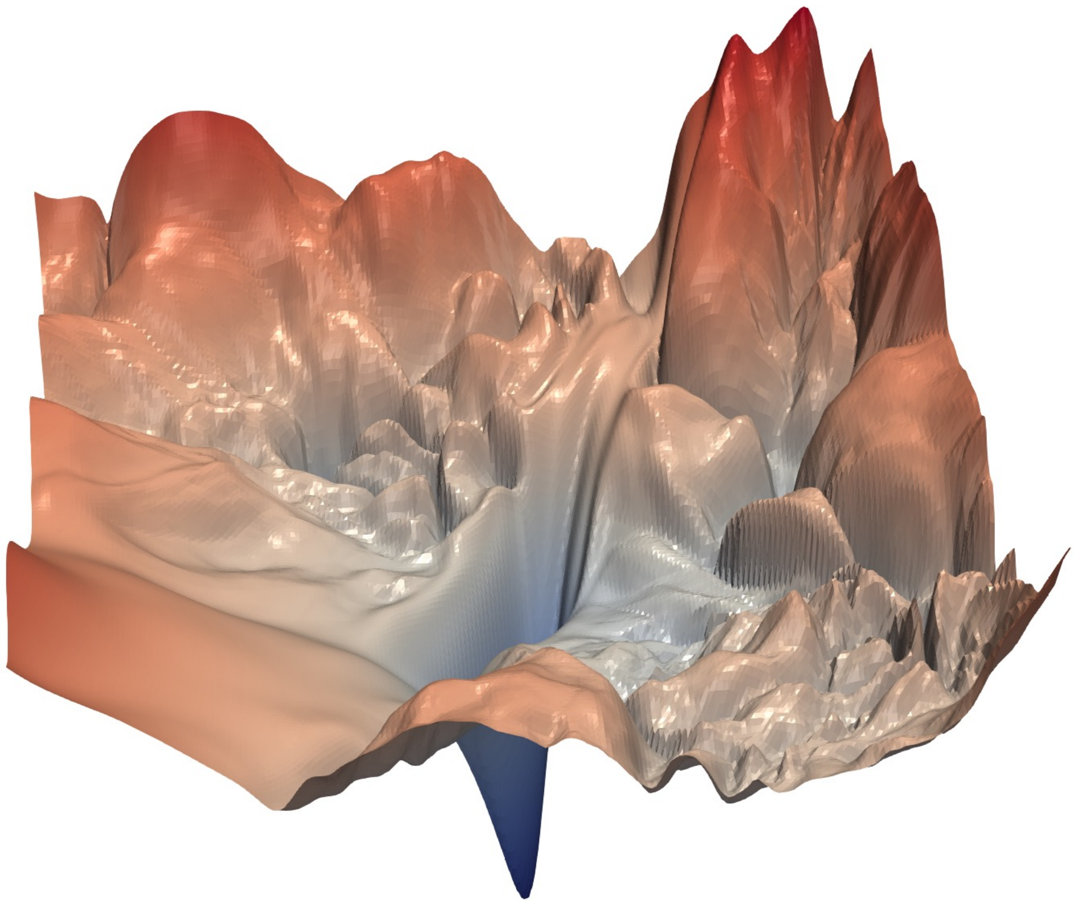
Backward Pass with Numerical Optimization

- Calculate Error
- Stochastic Gradient Descent
- Go towards minimum
- Correct network with chain rule
- Hopefully the global minimum



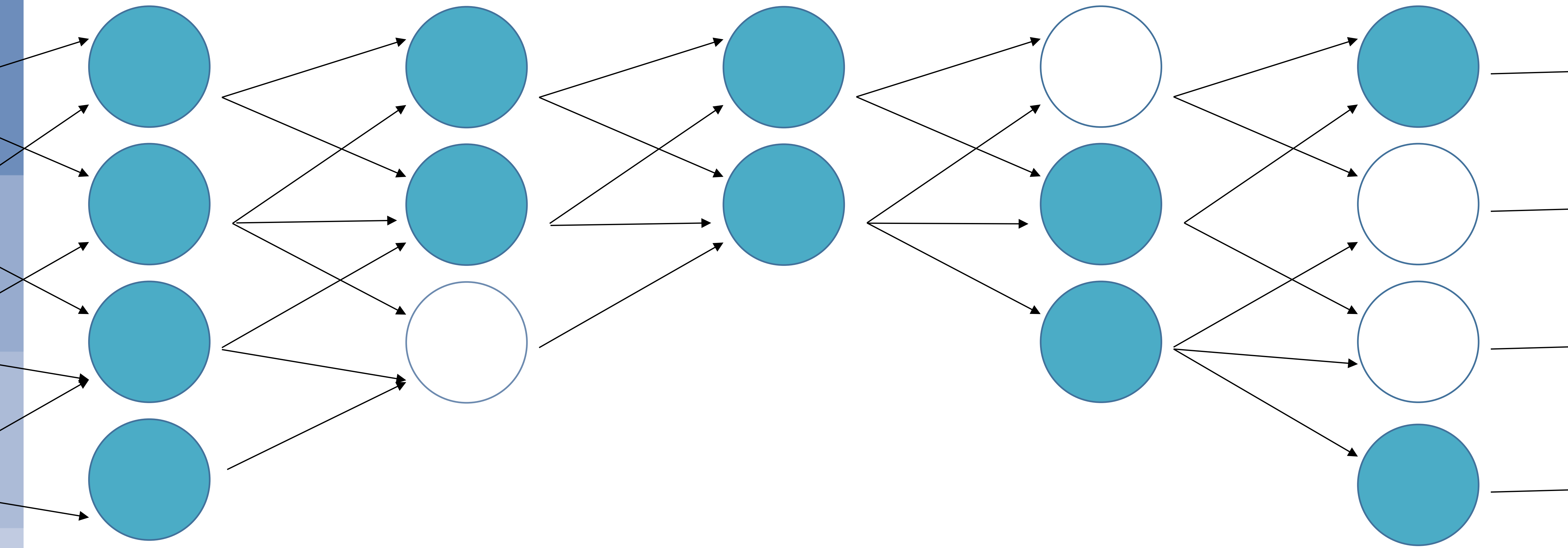
Realistic choices during training

- Loss surface usually highly irregular
- Different architectures choices change surface
- Take small steps toward minimum
- Use averaging and momentum
  - Adam optimiser
- Regularisation for better optimum



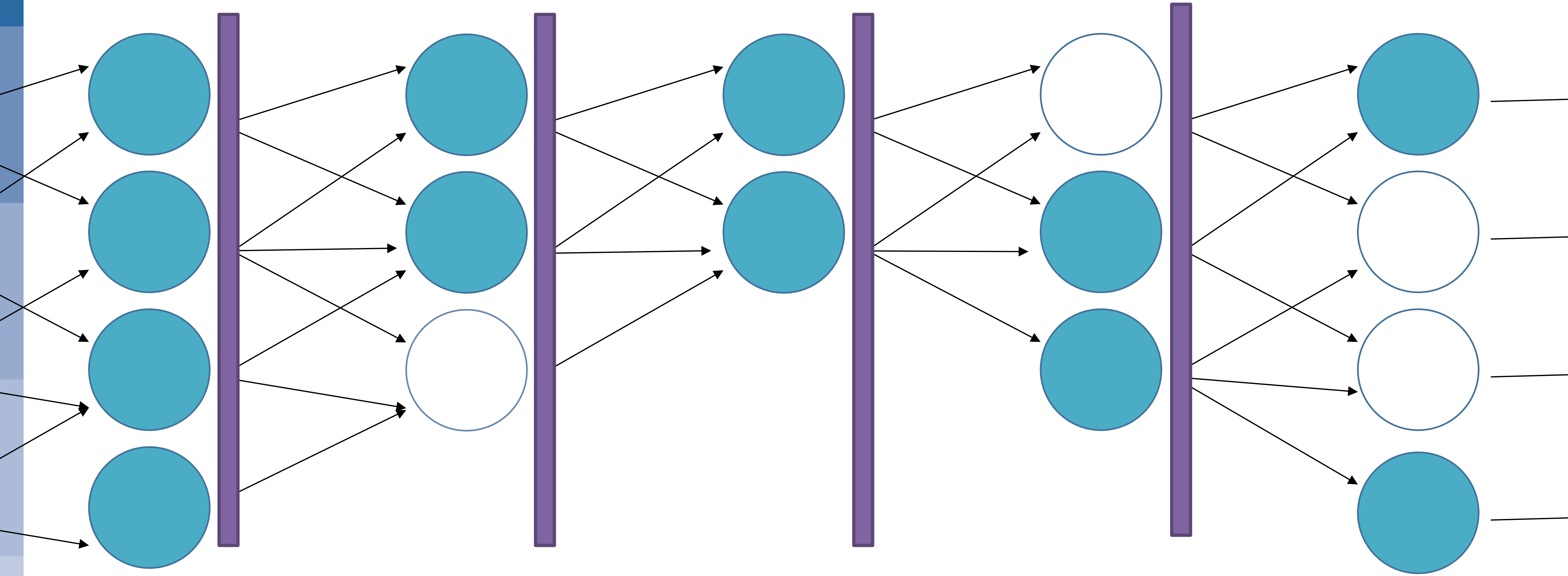
Li, et al. 2017: Visualizing the Loss Landscape of Neural Nets

Regularisation using Dropout



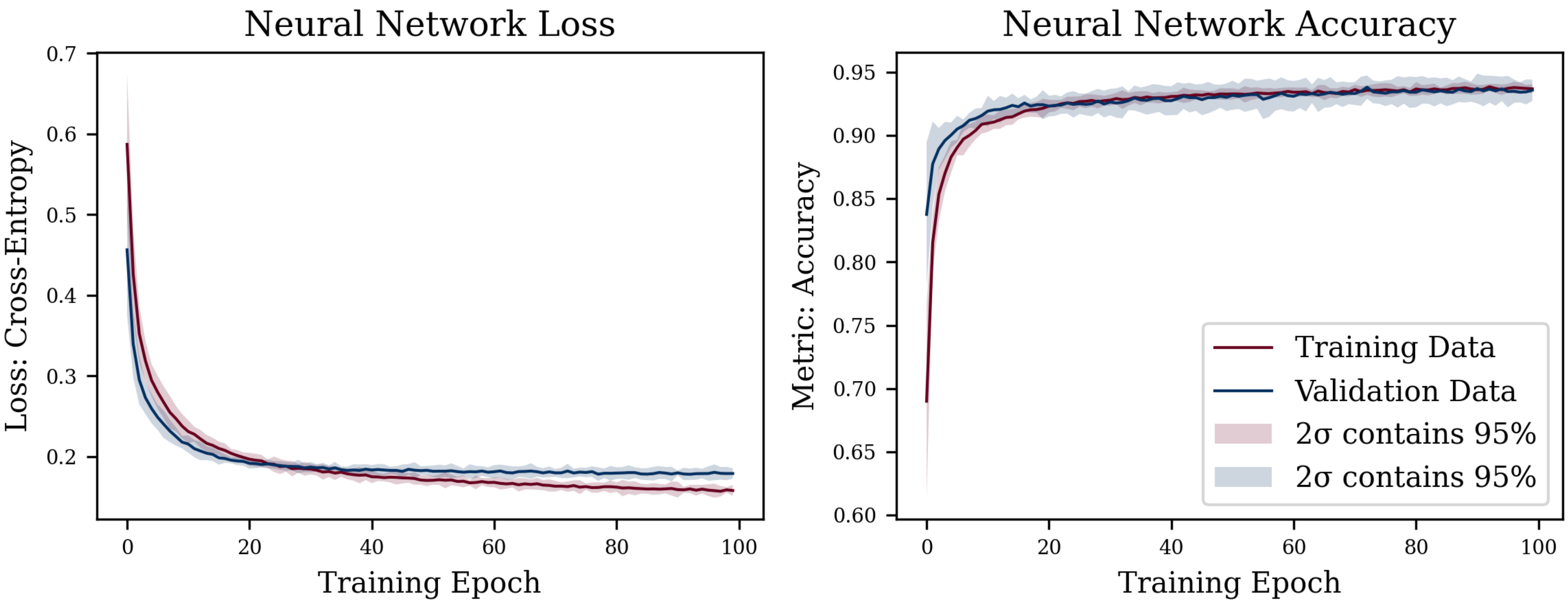
Neuron switched off

Standardisation using BatchNormalisation



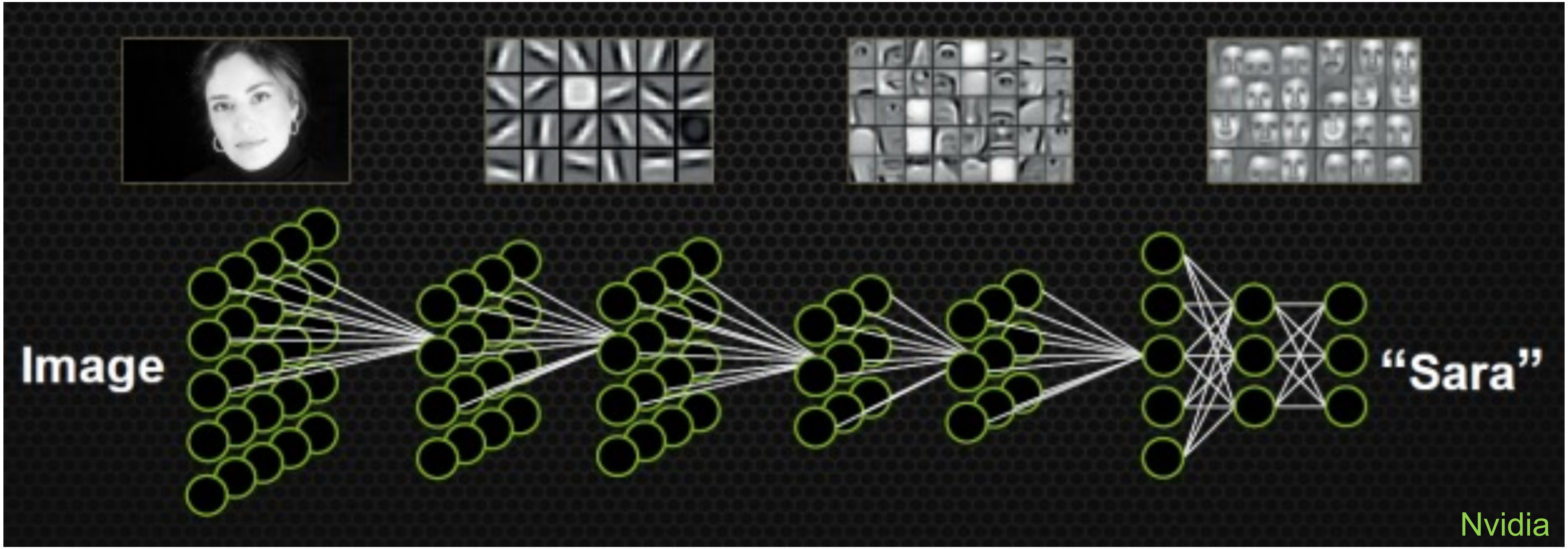
Normalise each batch of data

Loss and Accuracy Curves Converging

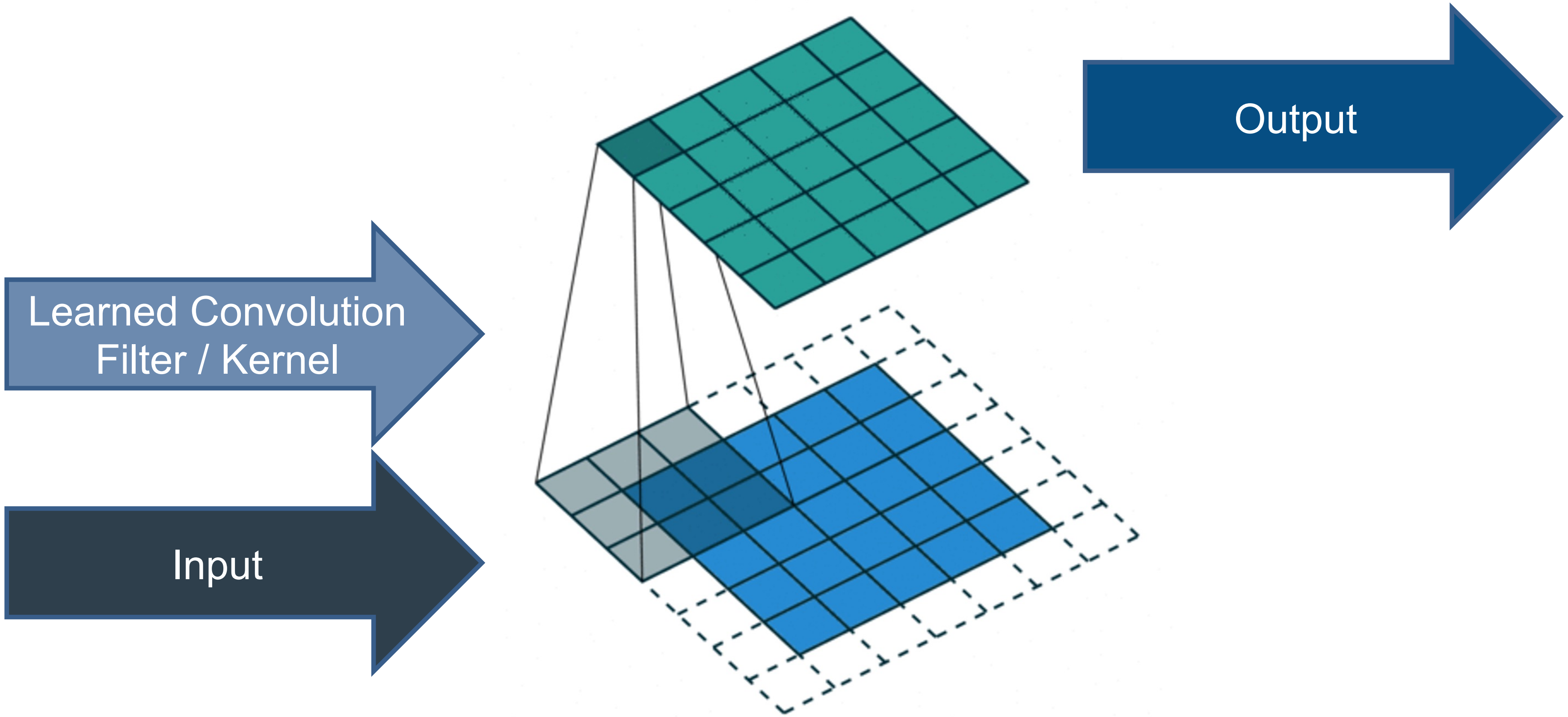


Working with Spatial Data

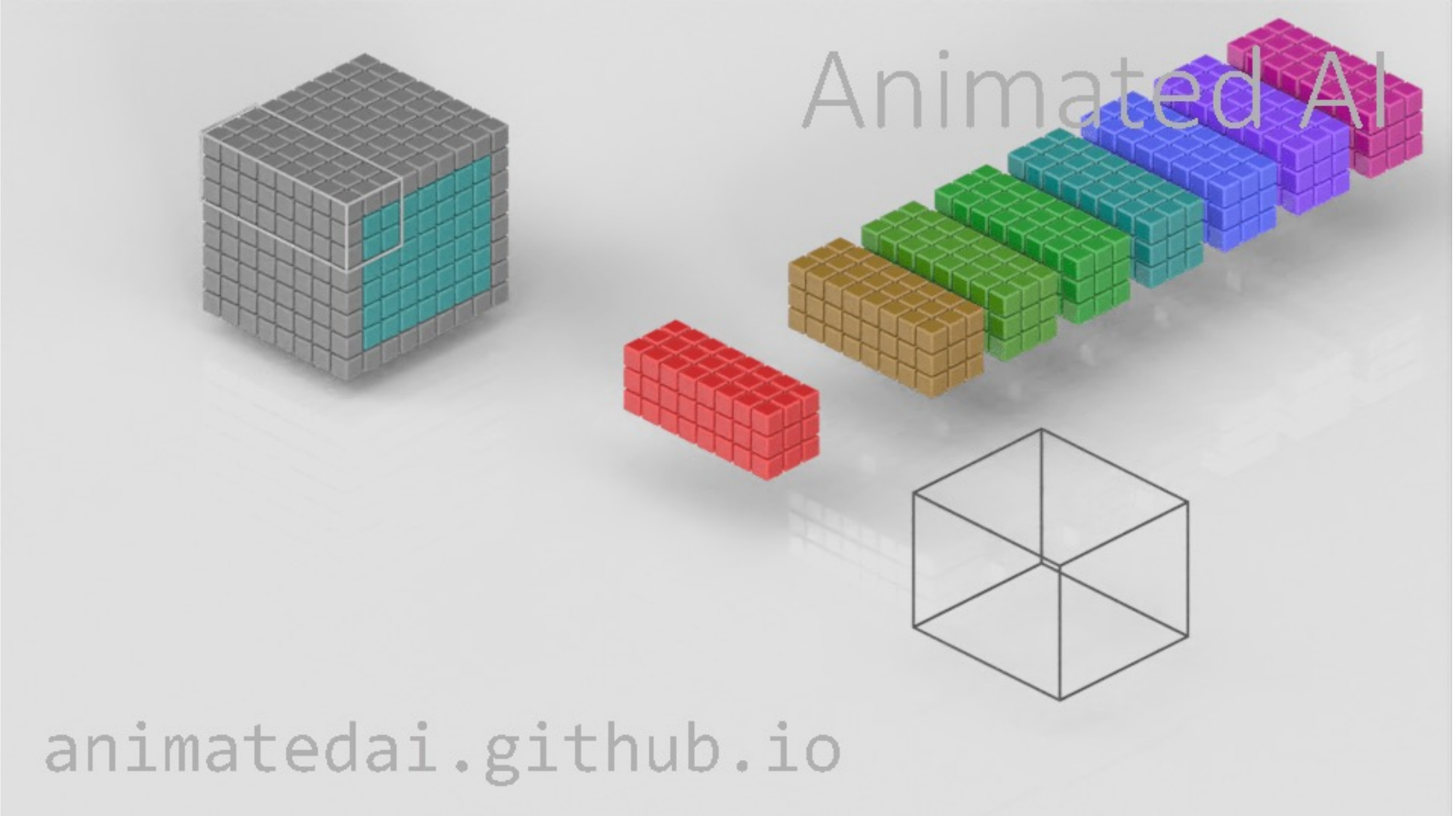
Networks on Images



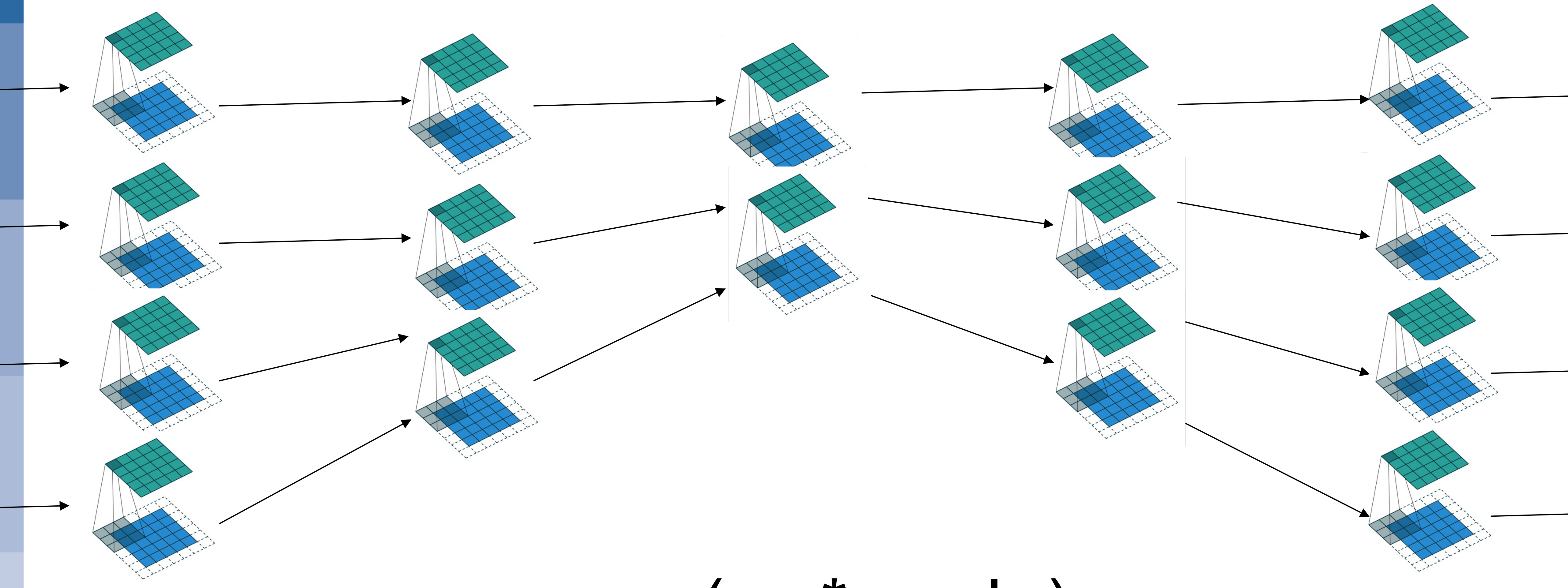
2D Convolutions



Multiple Convolutional Kernels in a Network Layer



Convolutional Neural Networks – Overly Simplified



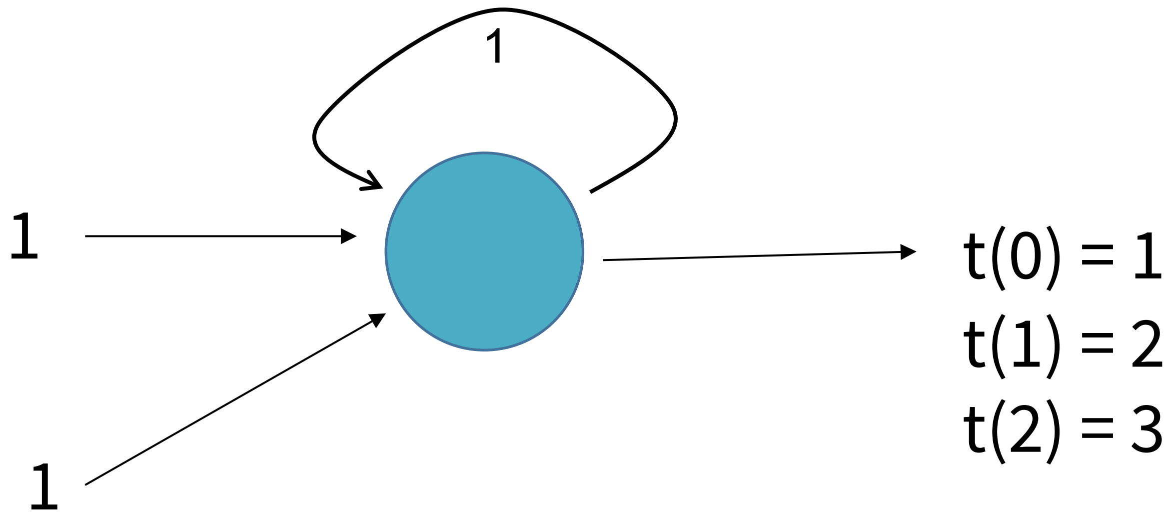
$$y = \sigma (\omega * x + b)$$

Convolutional Neural Networks

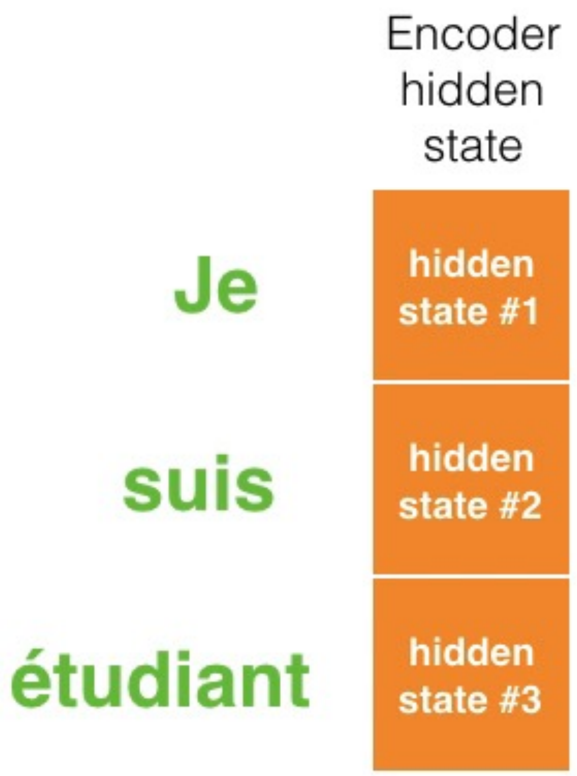
- Works with Locally Connected Data, e.g.
  - Photos
  - Satellite data
  - Weather fields
- Convolutional filters are learnt from data
- Compression changes focus of different layers
- Convolutions share weights and reduce computation

Working with Sequential Data

The simplest recurrent network

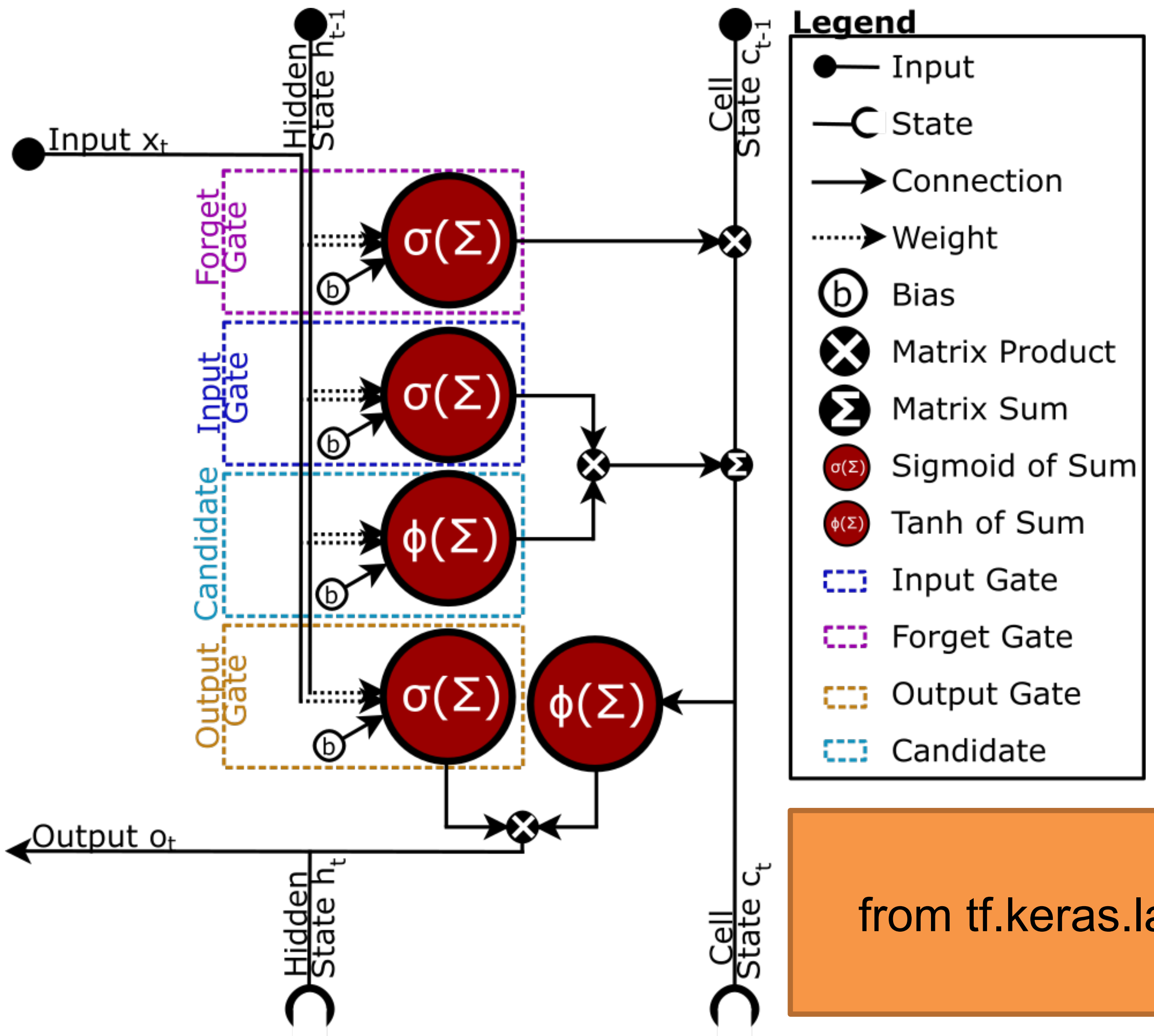


ECMWF



Encoder hidden state
Je
suis
étudiant
hidden state #1
hidden state #2
hidden state #3

Long Short-Term Memory (LSTM)



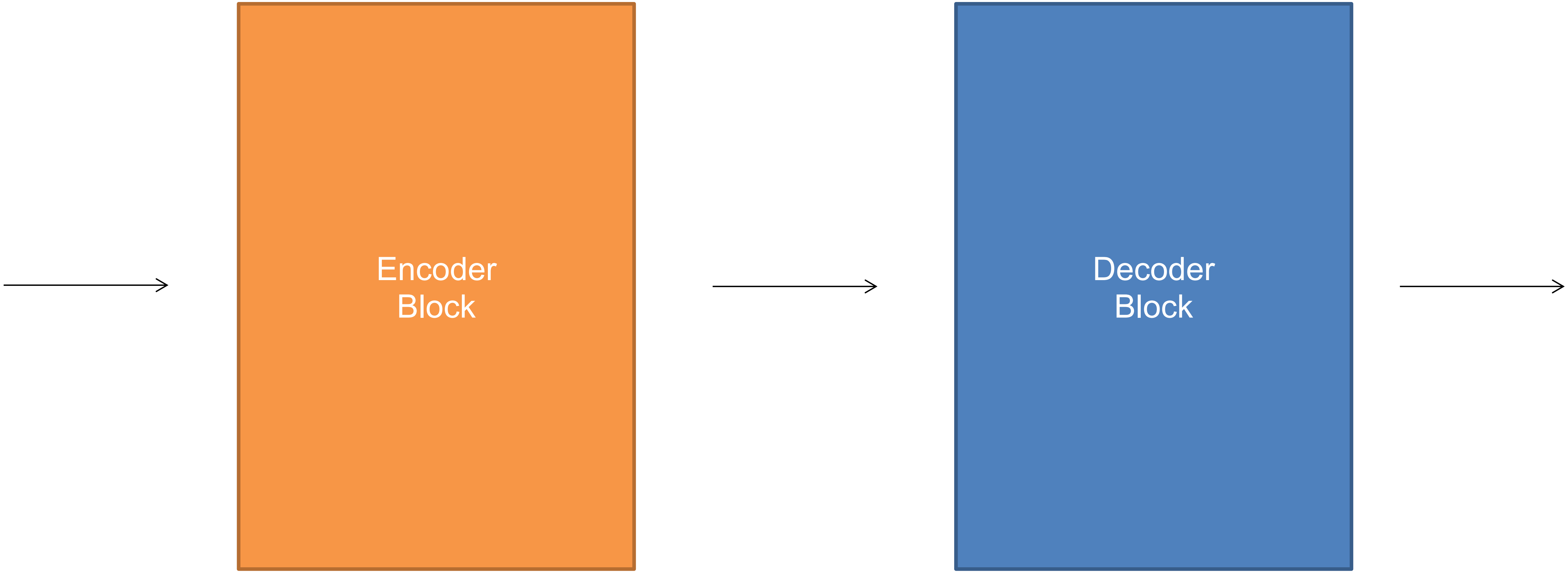
from tf.keras.layers import LSTM

Recurrent Neural Networks

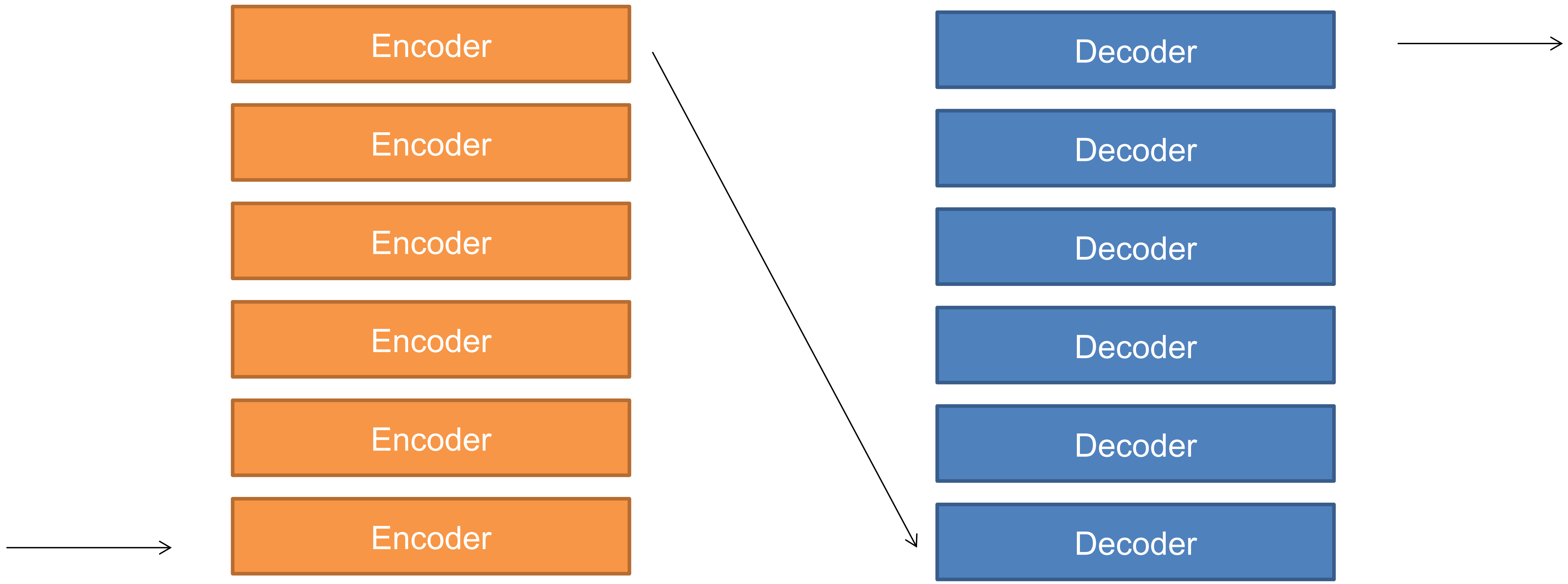
- Work with Sequences, e.g.
  - Text
  - Time Series
- Contain Feedback loop
- LSTM Cells contain a state from data
- Context for prediction is limited

Transformers

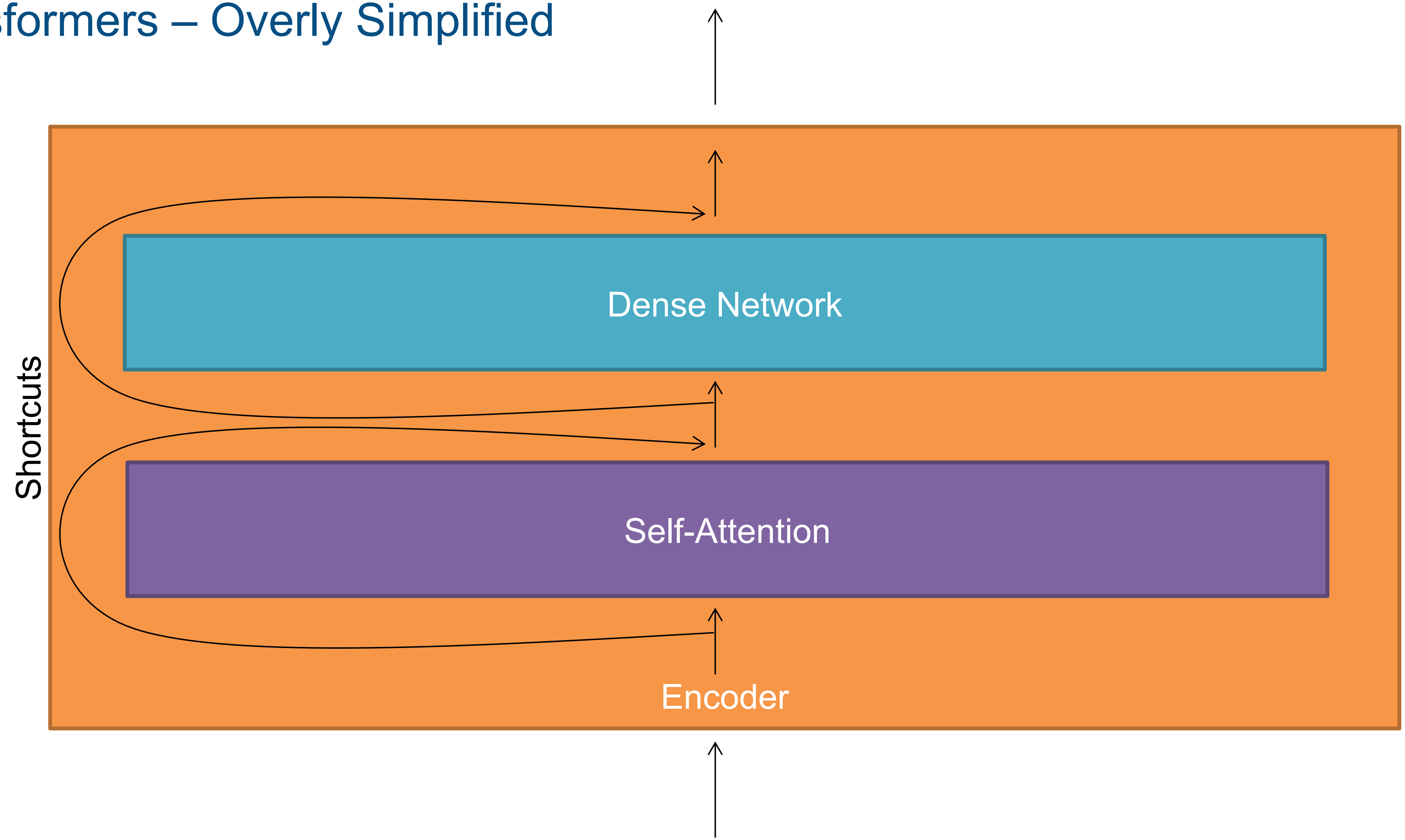
Transformers – Overly Simplified



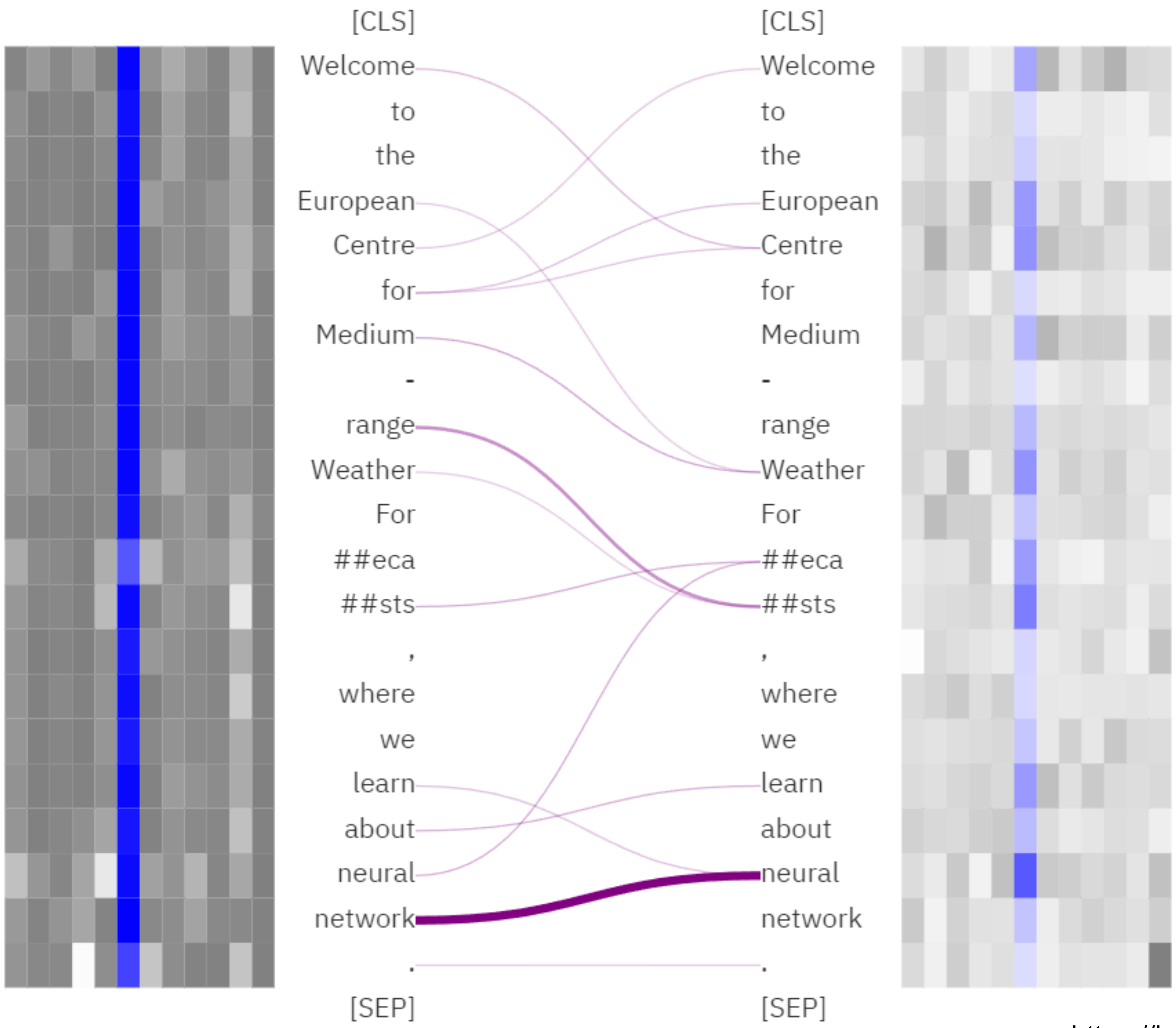
Transformers – Overly Simplified



Transformers – Overly Simplified

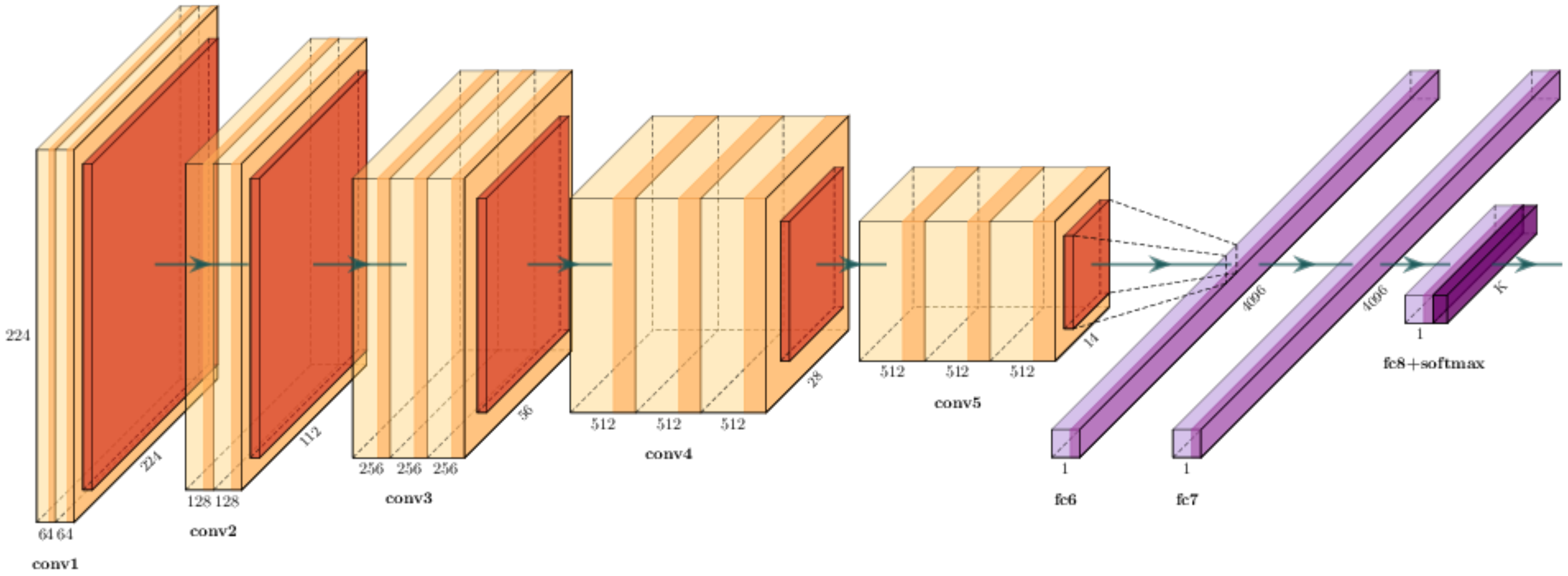


Self-Attention working on a sentence

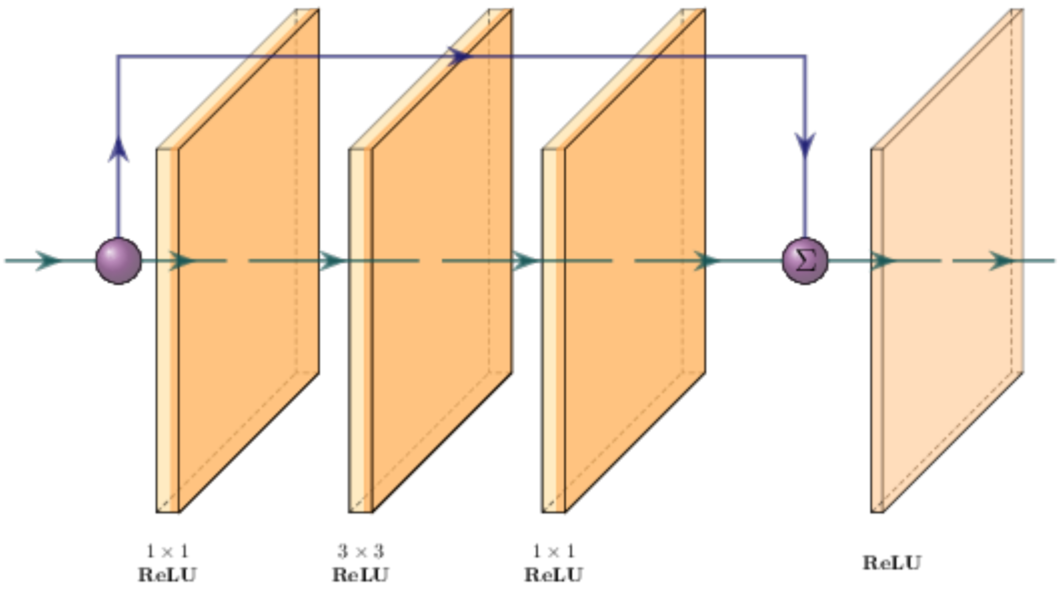


Combining Concepts into Architectures

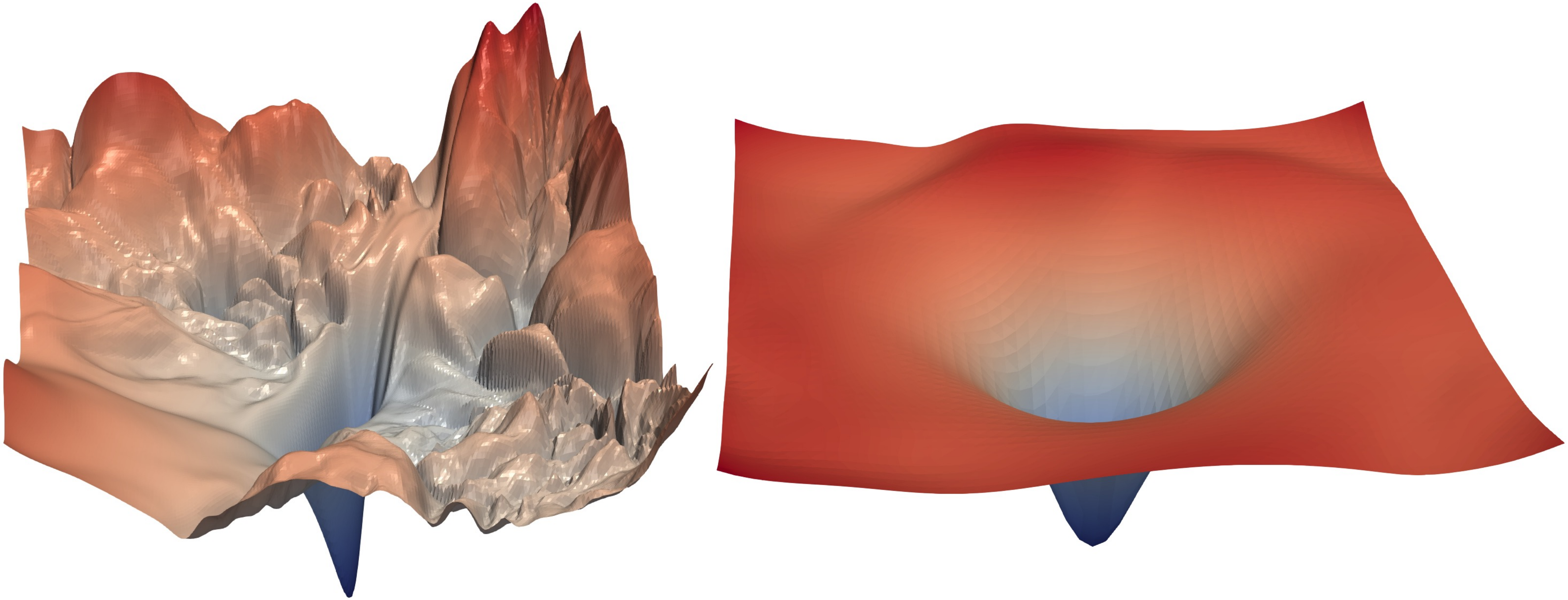
CNN + Dense: Classification Architecture (VGGNet-16)



ResNet Blocks: Utilizing Shortcuts

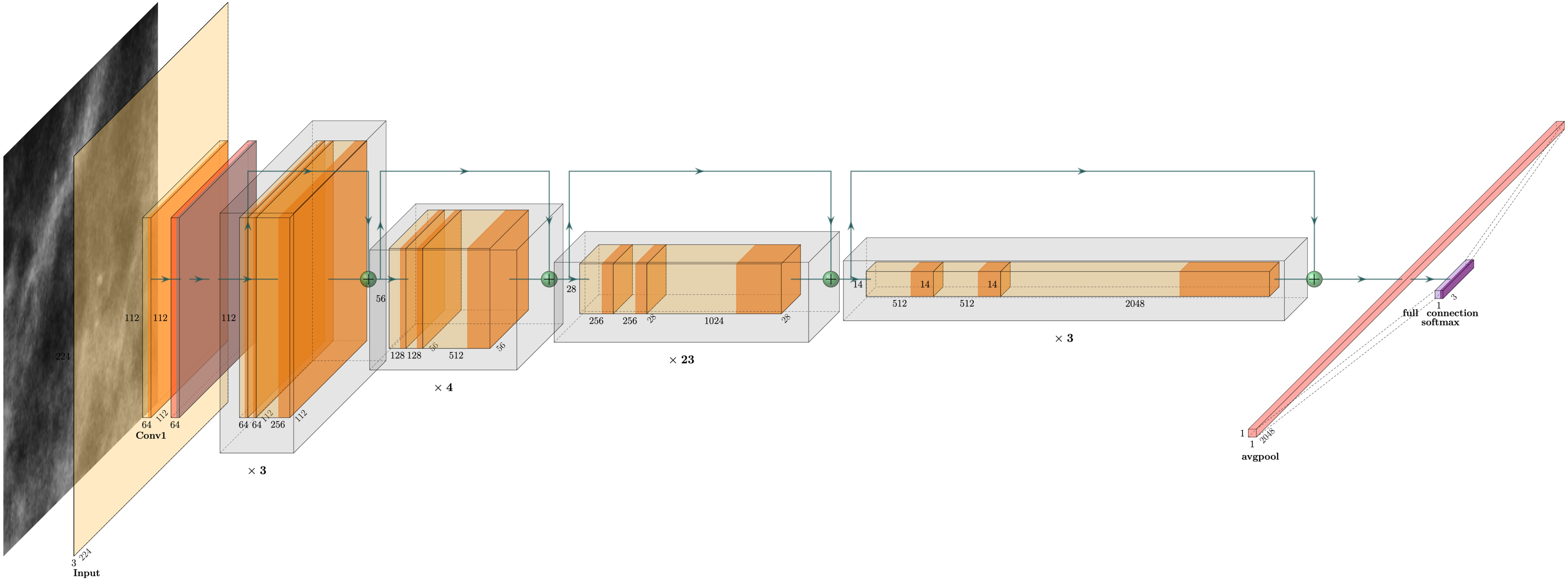


Why we use Residual Connections

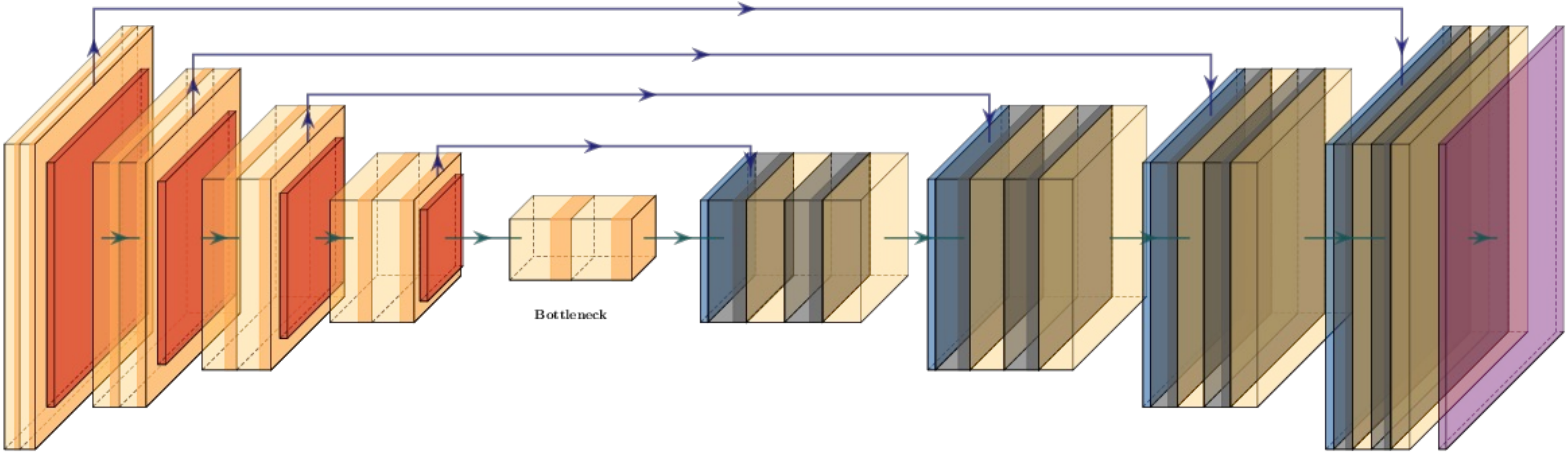


Li, et al. 2017: Visualizing the Loss Landscape of Neural Nets

Going deep: ResNet-101



Unet: Utilizing Compression for Encoding / Decoding



Encoder

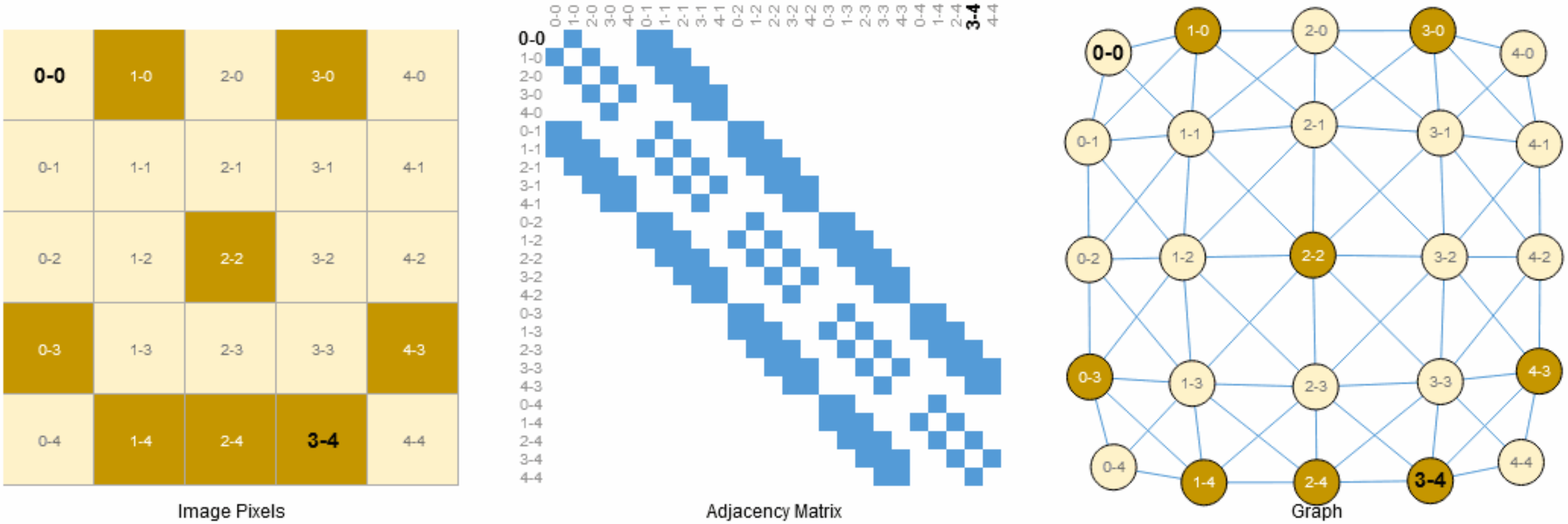
Bottleneck

Decoder

Why we use Compression / Latent Spaces

Graph Neural Networks

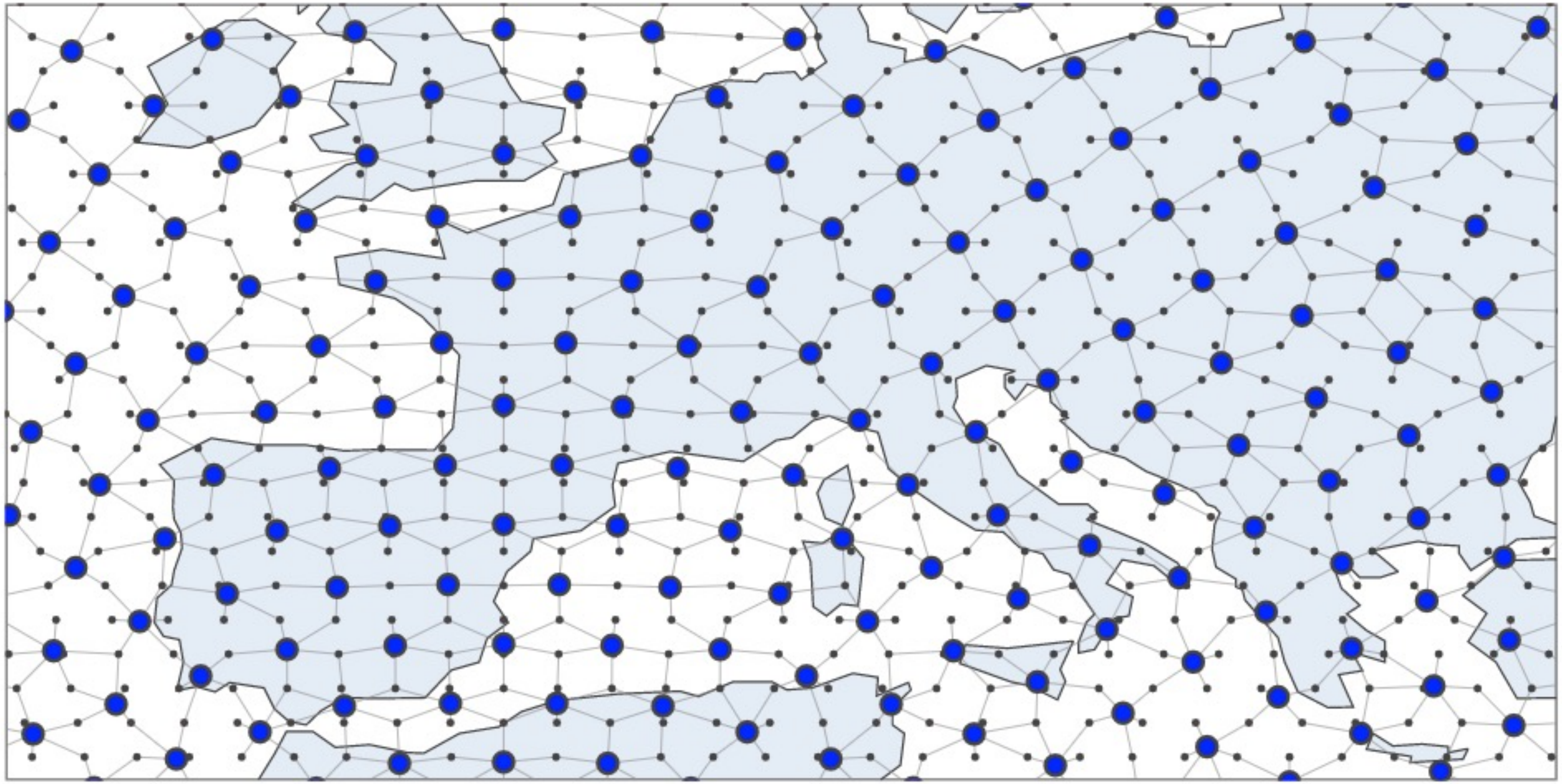
Defining Operations on Graphs



Click on an image pixel to toggle its value, and see how the graph representation changes.

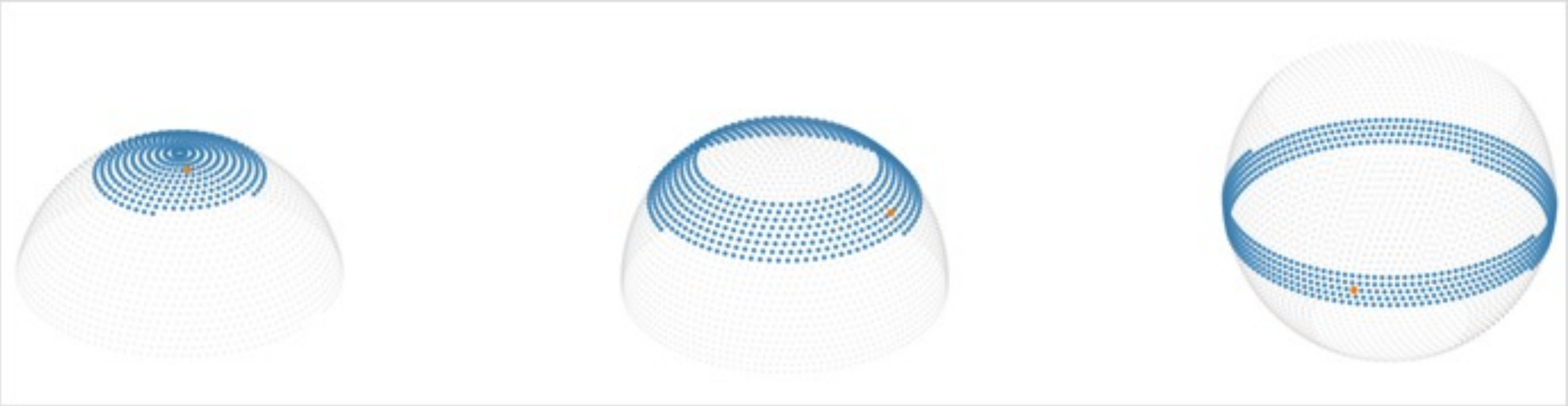
 Sanchez-Lengeling, et al., "A Gentle Introduction to Graph Neural Networks", Distill, 2021.

Defining Operations on Graphs: Convolutions



AIFS v0.1

Defining Operations on Graphs: Transformers



AIFS v0.21

https://www.ecmwf.int/en/about/media-centre/aifs-blog/2024/first-update-aifs

Conclusion

What We Learned

- Neural Network Training
- Network Types
  - Dense Neural Networks
  - Convolutional Neural Networks
  - Recurrent Neural Networks
  - Transformers
  - Graph Neural Networks
- Example Architectures
- Compression
- Shortcuts / Residual Connections