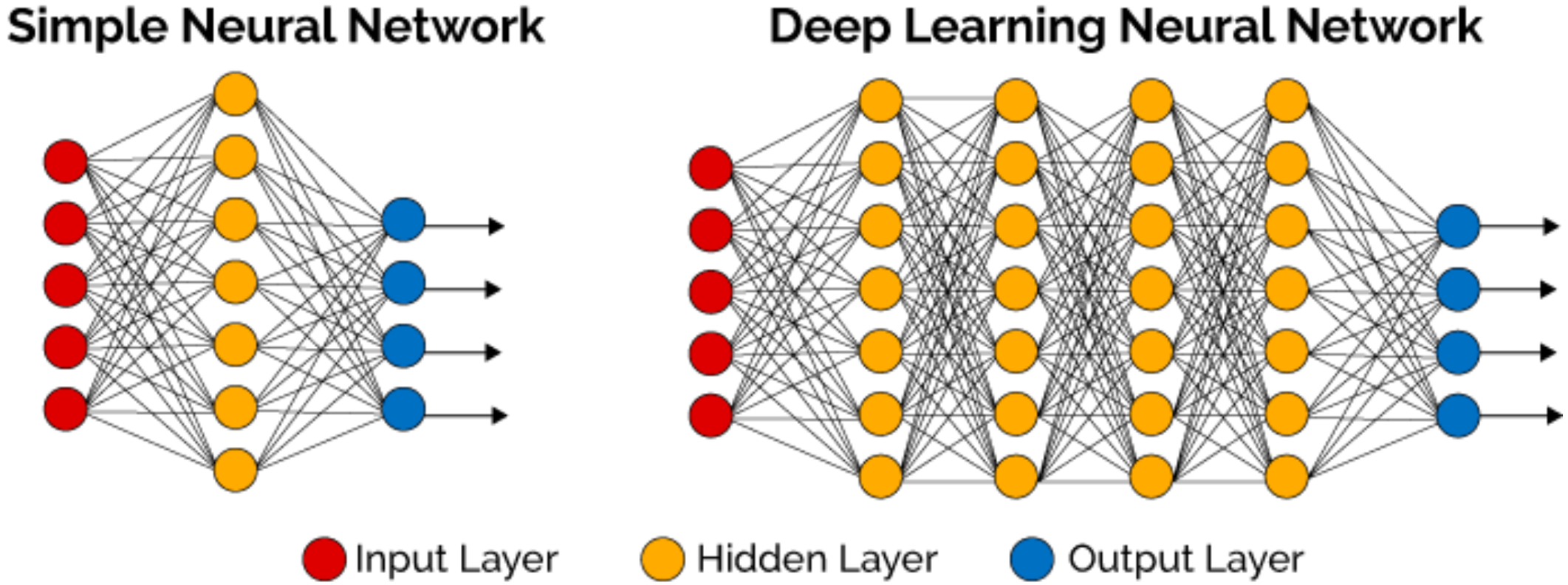
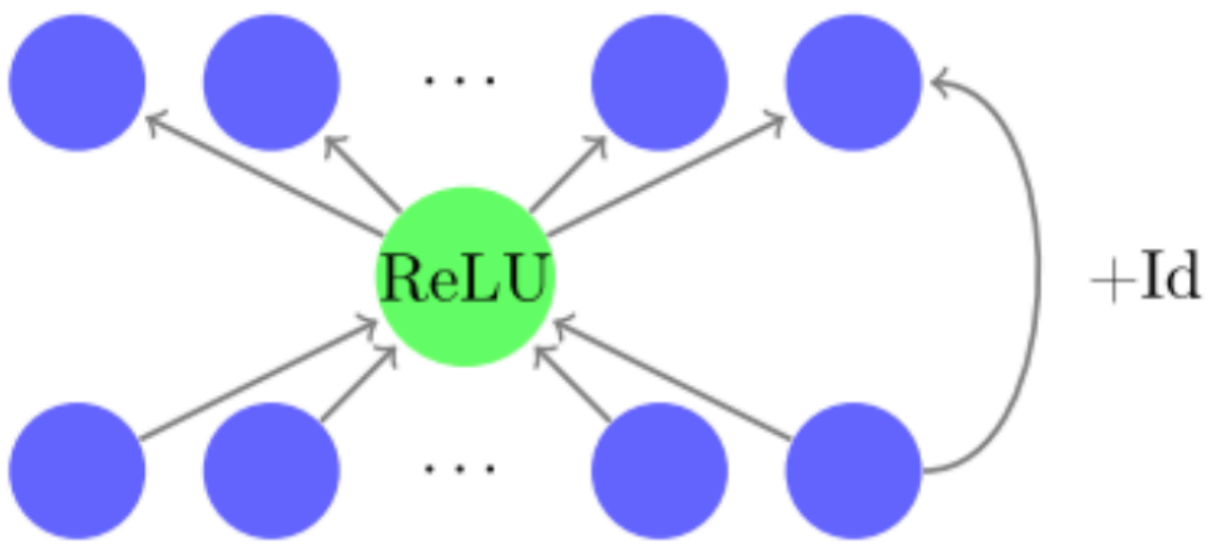
Introduction To Various Neural Network Architectures

Ankur Bohra

What are Neural Networks?

Universal Approximation Theorem





Universal Approximation Theorem for The One-Neuron ResNet

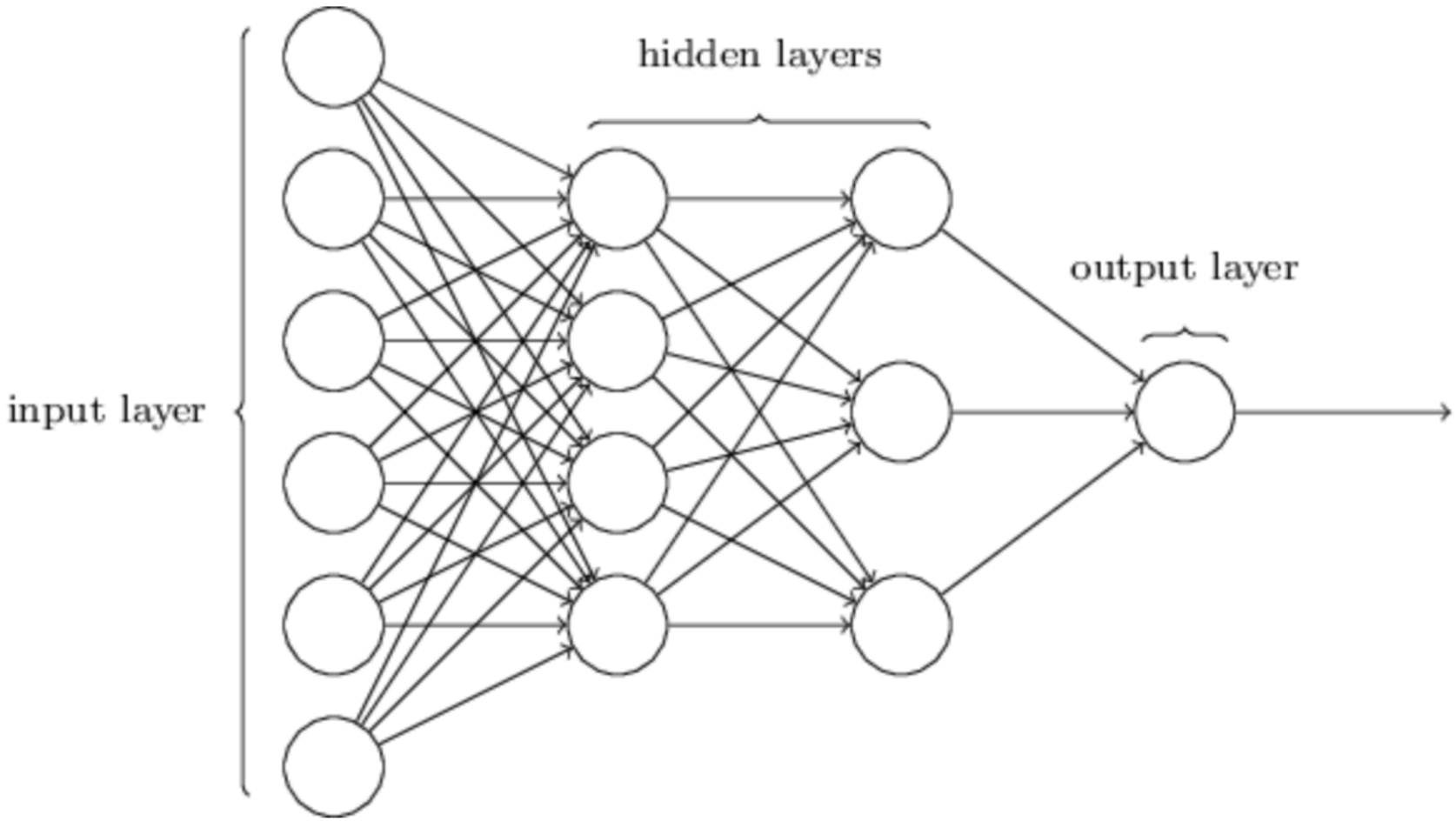
Why do you need different architectures?

Why?

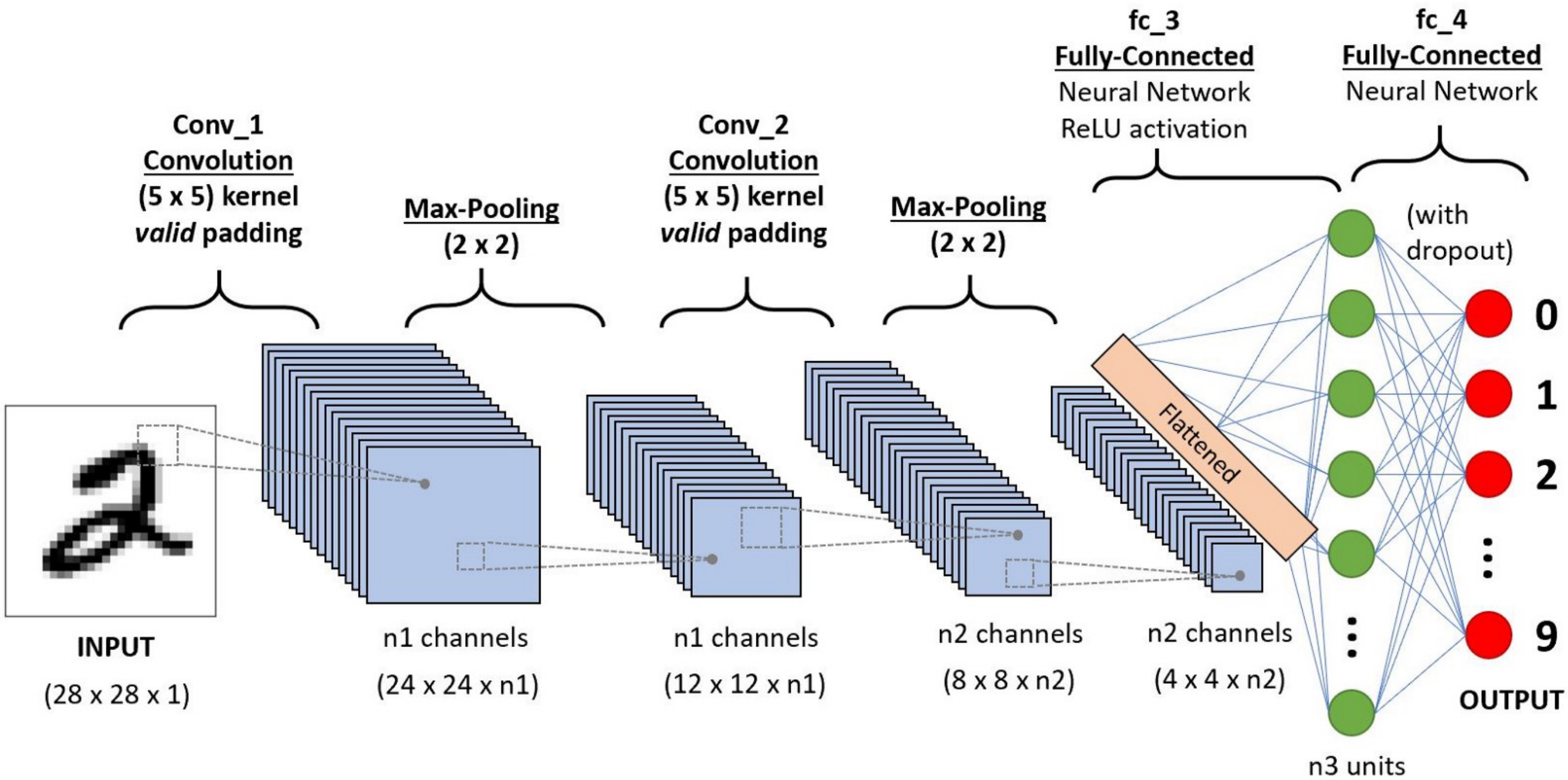
- Convergence
- Task
- Input/output
- …

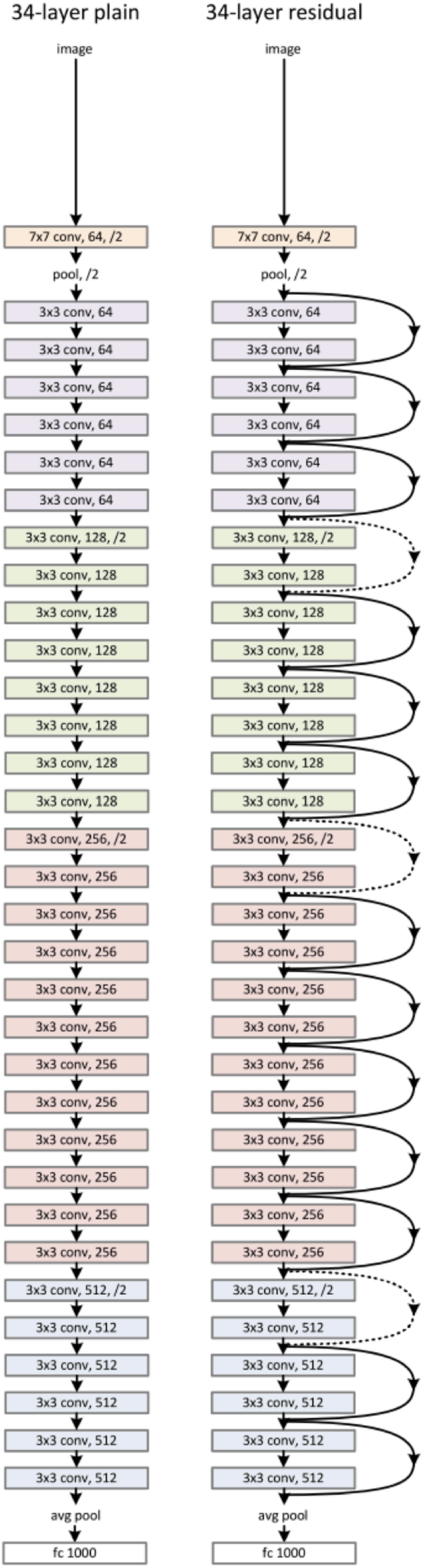
Cat

Feed Forward Network



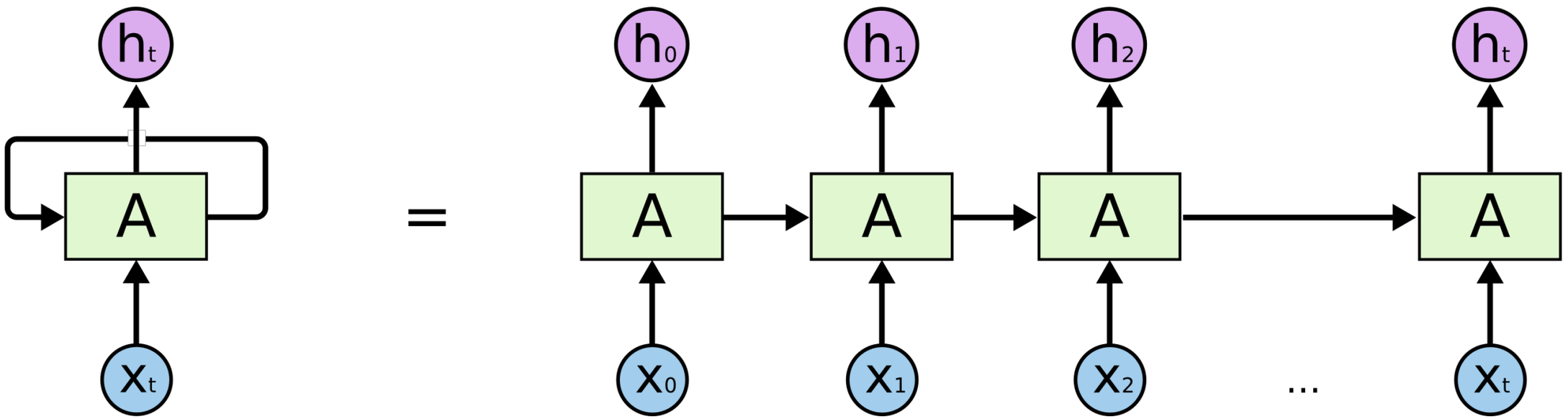
CNN


Conv_1
Convolution
(5 x 5) kernel
valid padding
Max-Pooling
(2 x 2)
Conv_2
Convolution
(5 x 5) kernel
valid padding
Max-Pooling
(2 x 2)
fc_3
Fully-Connected
Neural Network
ReLU activation
fc_4
Fully-Connected
Neural Network
(with
dropout)
Flattened
INPUT
(28 x 28 x 1)
n1 channels
(24 x 24 x n1)
n1 channels
(12 x 12 x n1)
n2 channels
(8 x 8 x n2)
n2 channels
(4 x 4 x n2)
n3 units
0
1
2
9
OUTPUT

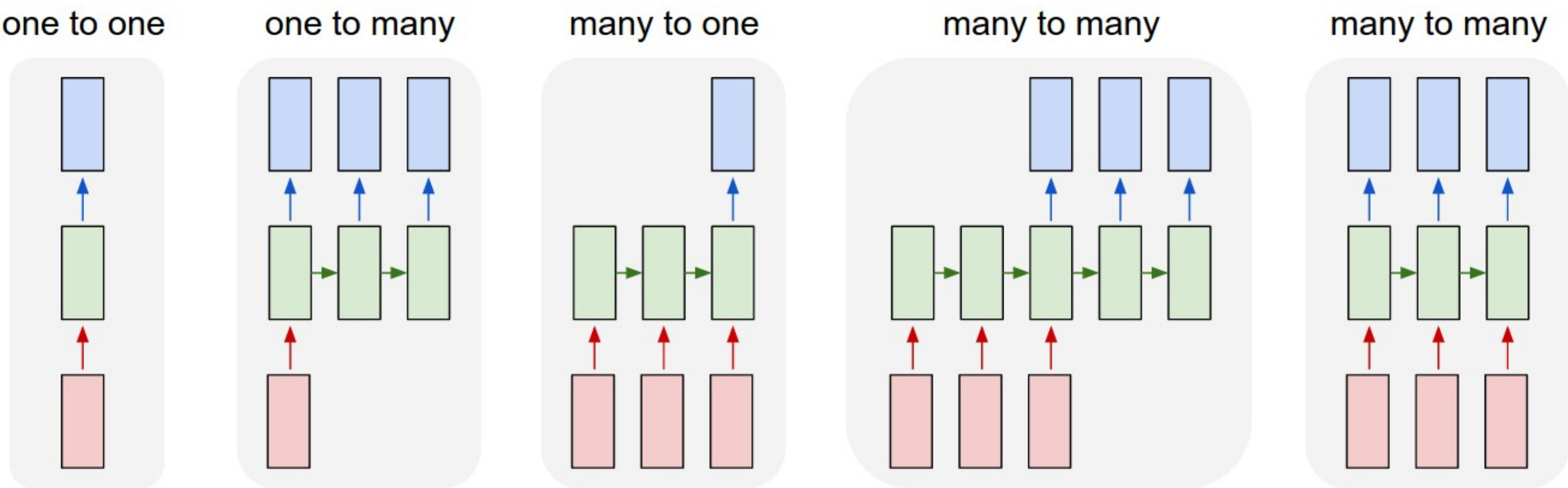

34-layer plain
34-layer residual
image
7x7 conv, 64, /2
pool, /2
3x3 conv, 64
3x3 conv, 128, /2
3x3 conv, 128
3x3 conv, 256, /2
3x3 conv, 256
3x3 conv, 512, /2
3x3 conv, 512
avg pool
fc 1000

Resnet

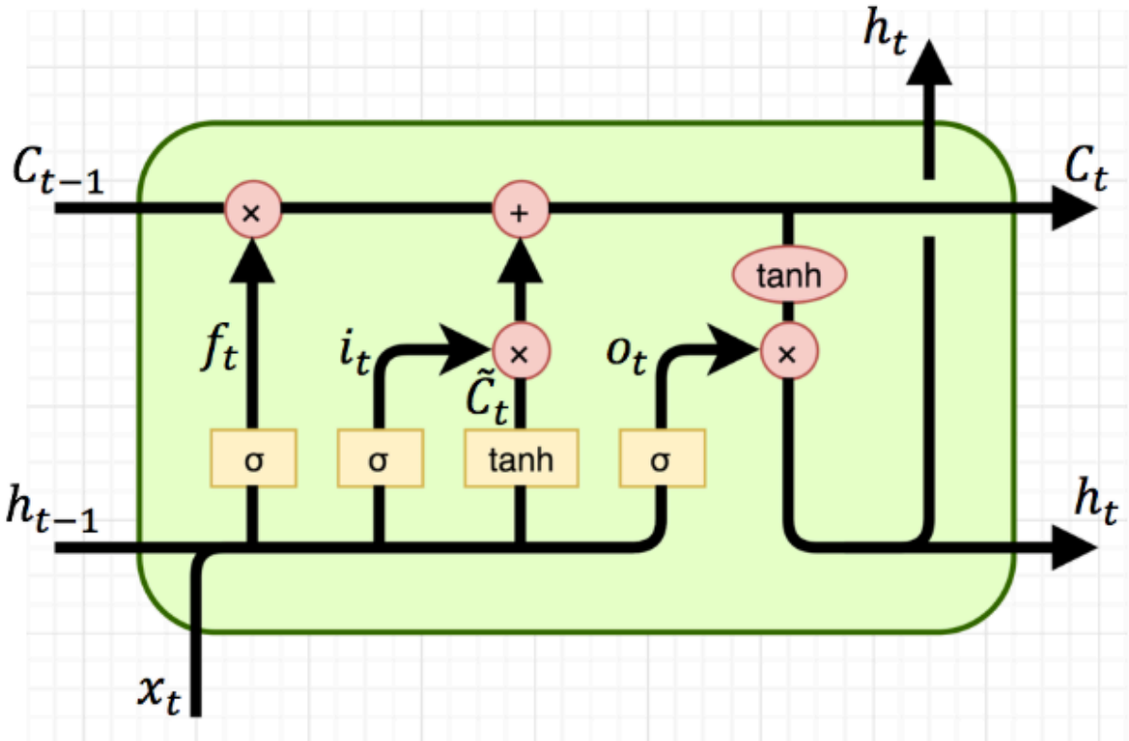
RNN



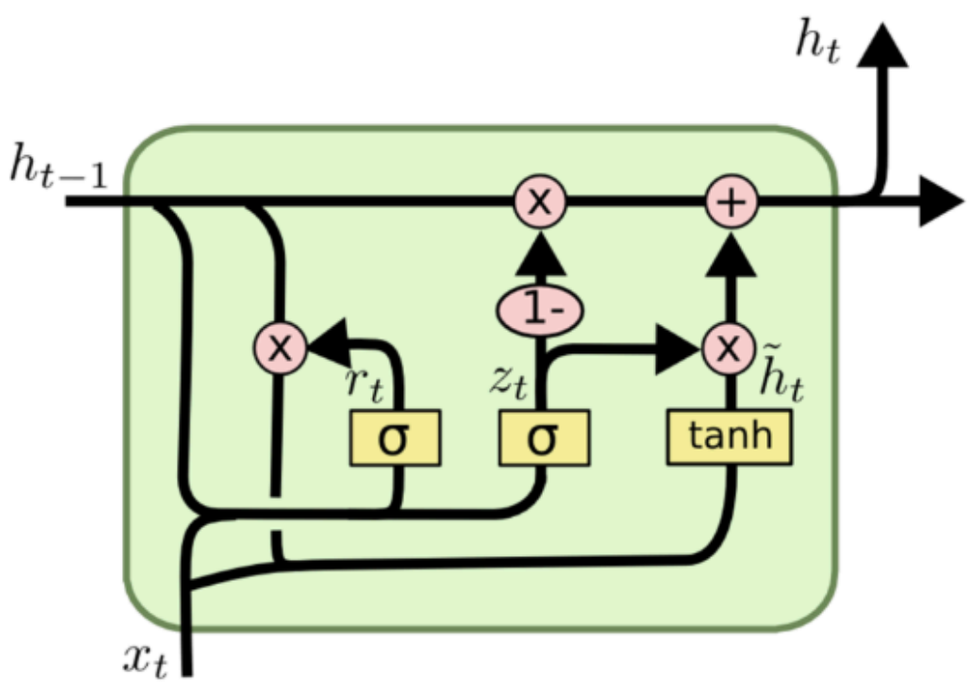
RNN Input/Output Sequence



LSTM and GRU

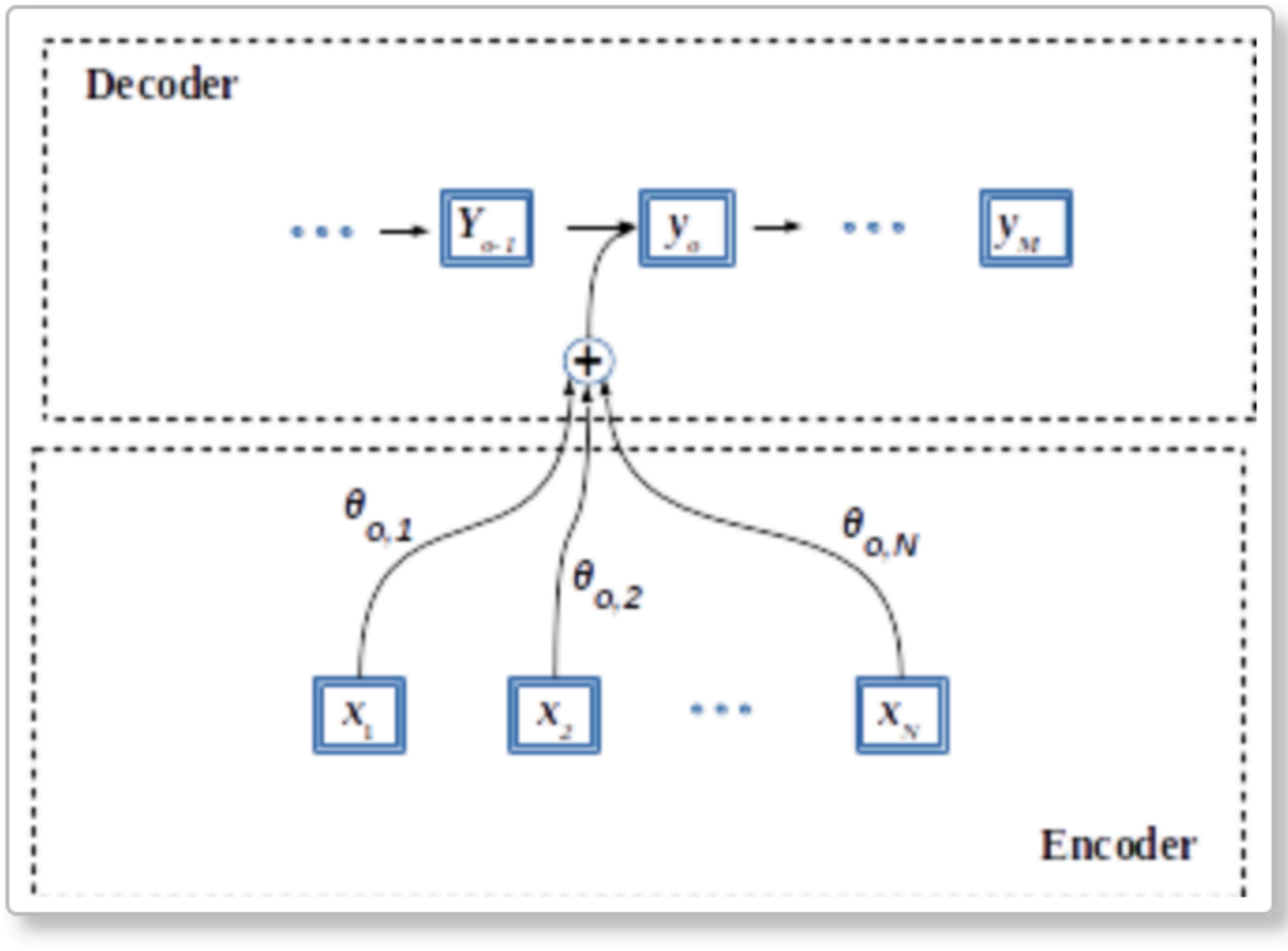


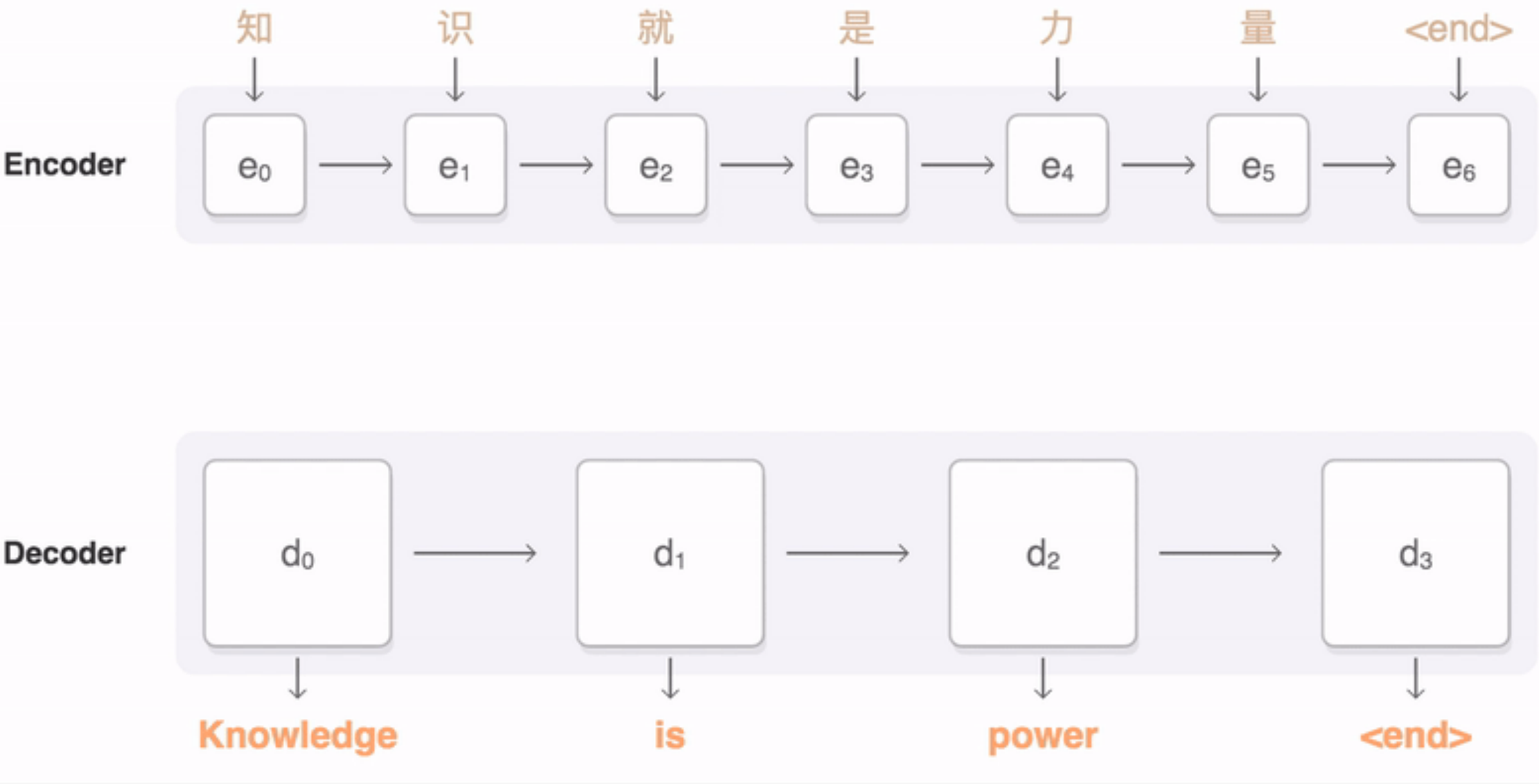
(a) Long Short-Term Memory



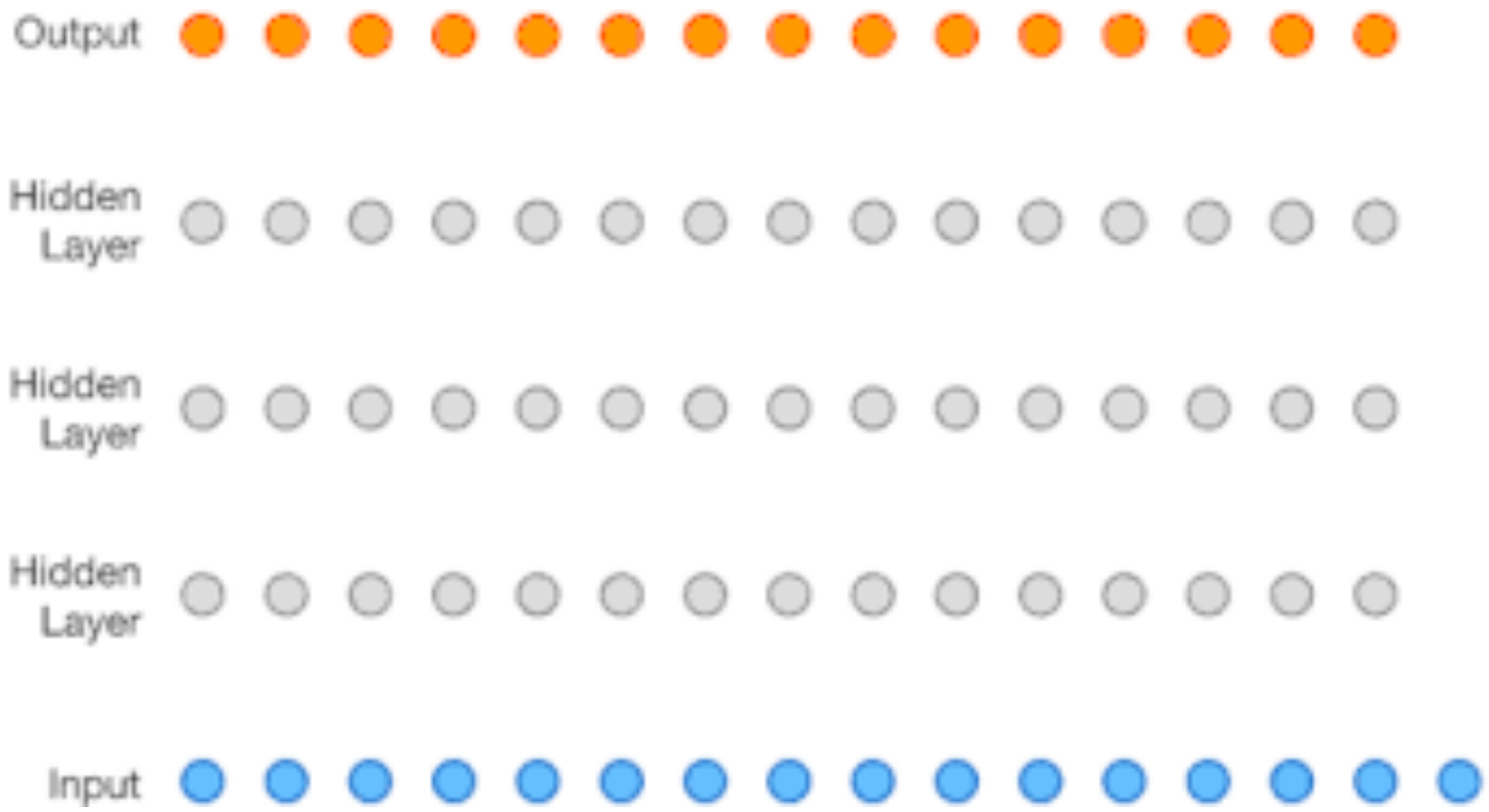
(b) Gated Recurrent Unit

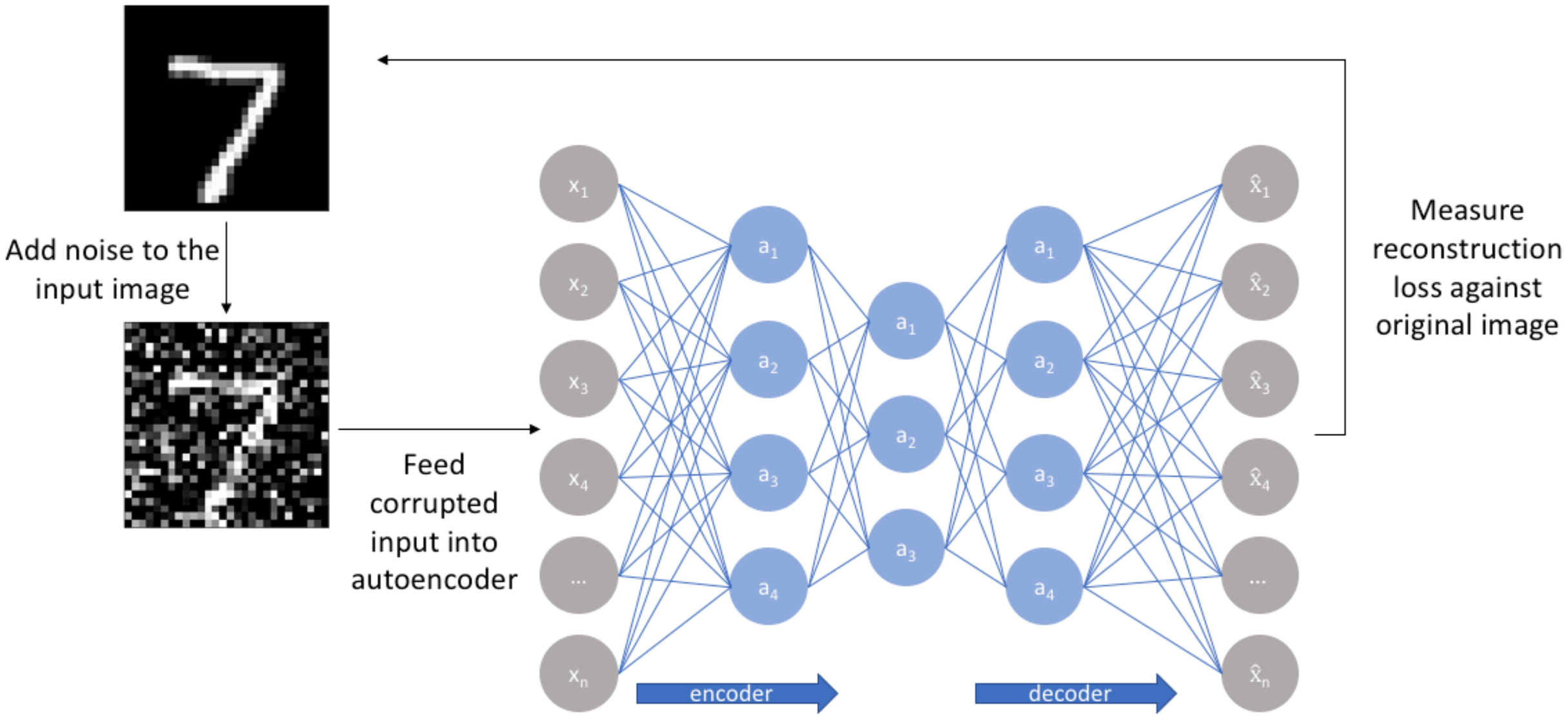
Attention Network




知
识
就
是
力
量
<end>
Encoder
e0
e1
e2
e3
e4
e5
e6
Decoder
d0
d1
d2
d3
Knowledge
is
power
<end>

Wavenet (Autoregressive)





Auto-encoders

Transformer

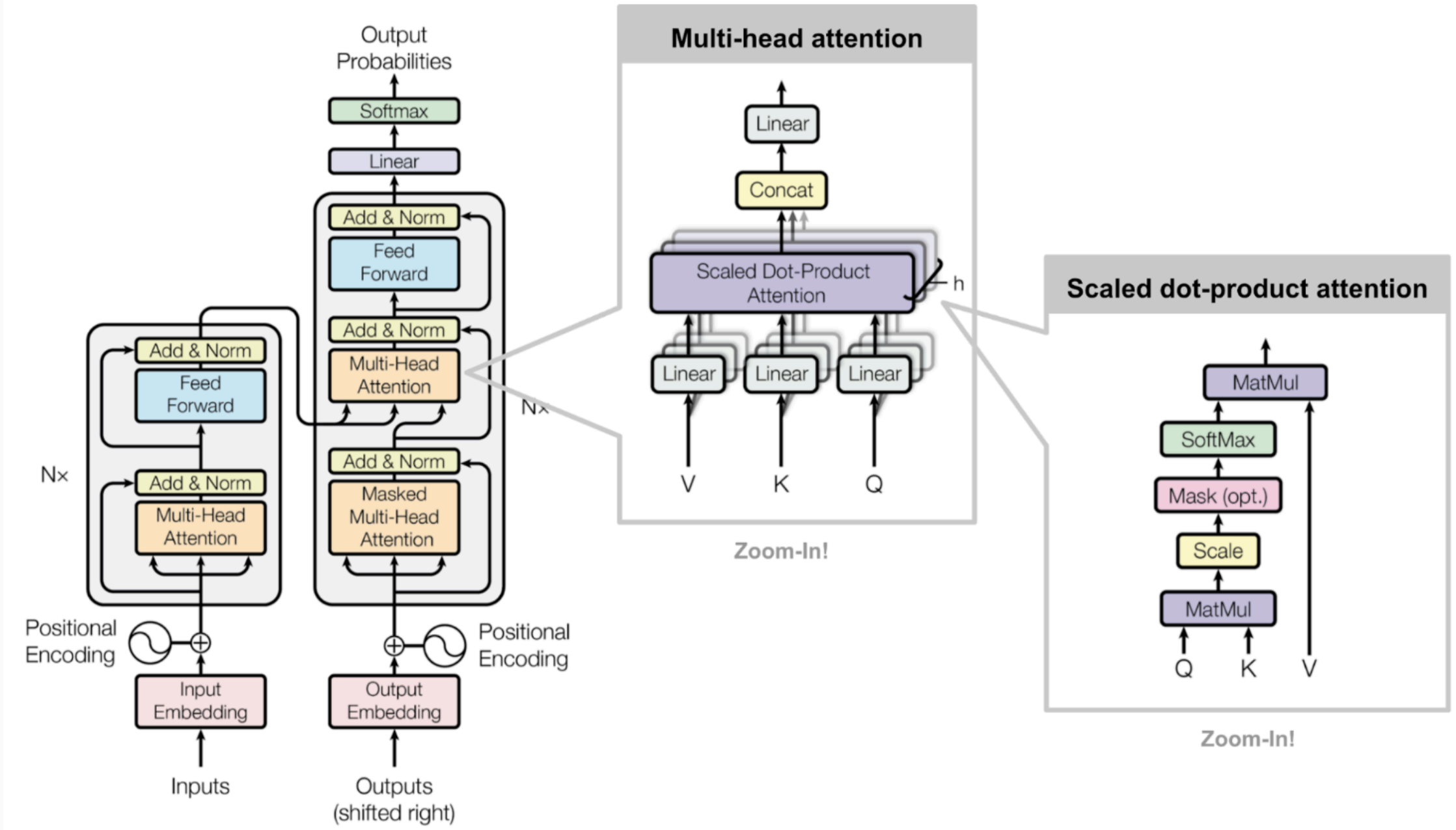
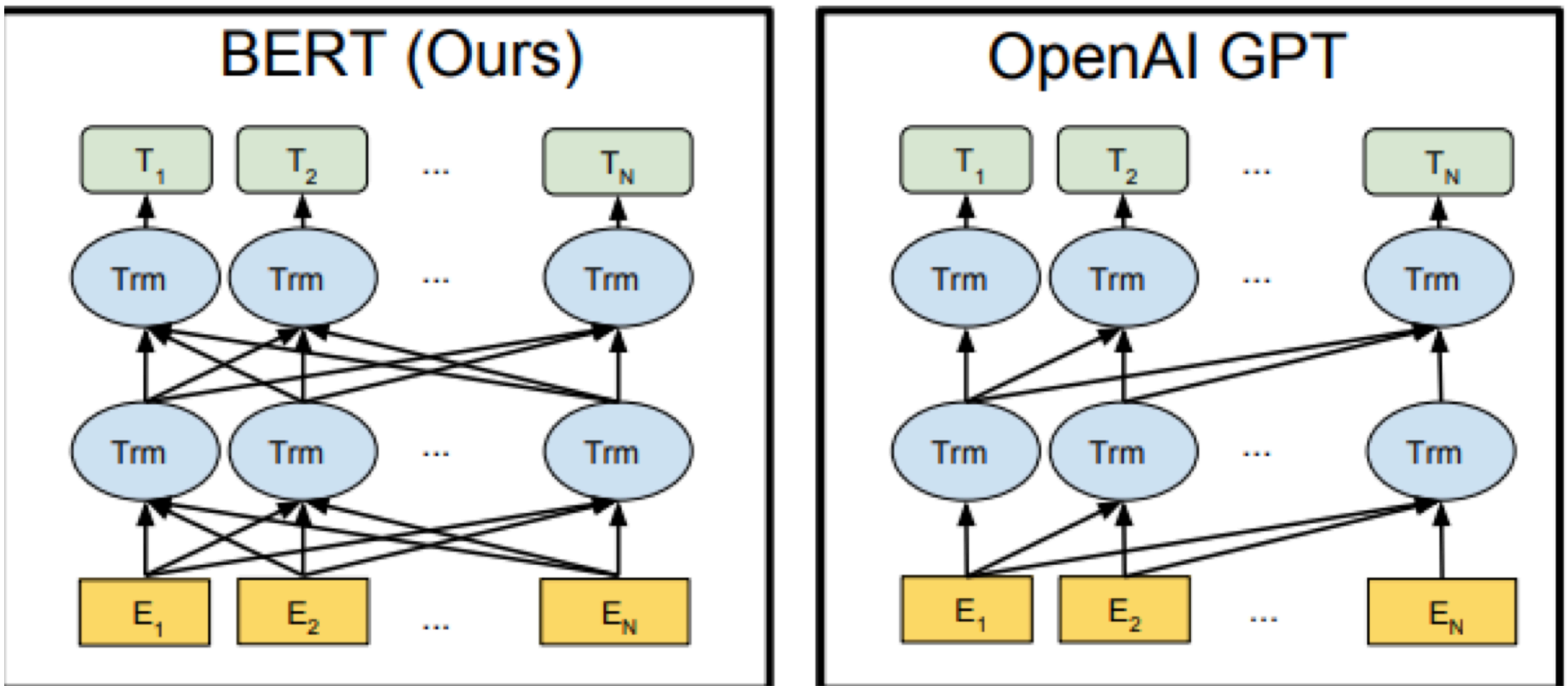
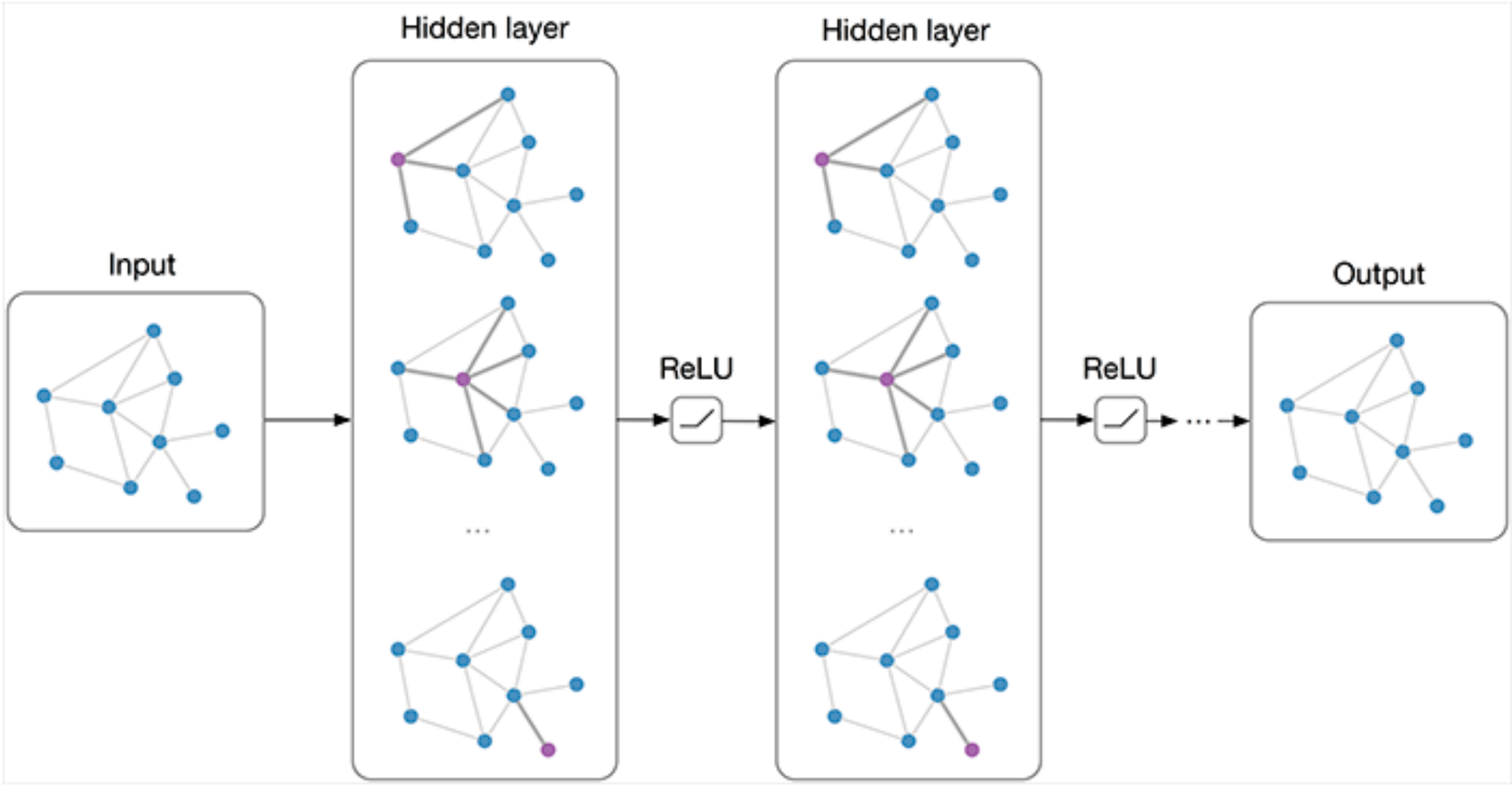


Fig. 17. The full model architecture of the transformer. (Image source: Fig 1 & 2 in Vaswani, et al., 2017.)

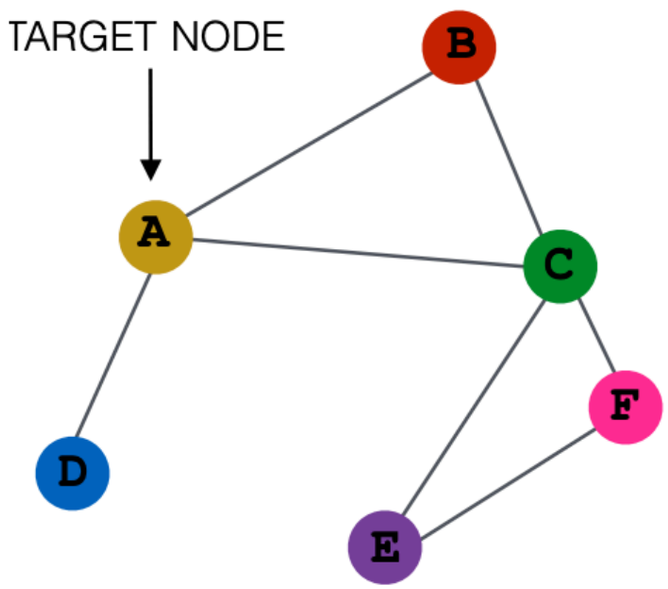
BERT & GPT


BERT (Ours)
T1
T2
...
TN
Trm
Trm
...
Trm
Trm
Trm
...
Trm
E1
E2
...
EN
OpenAI GPT
T1
T2
...
TN
Trm
Trm
...
Trm
Trm
Trm
...
Trm
E1
E2
...
EN

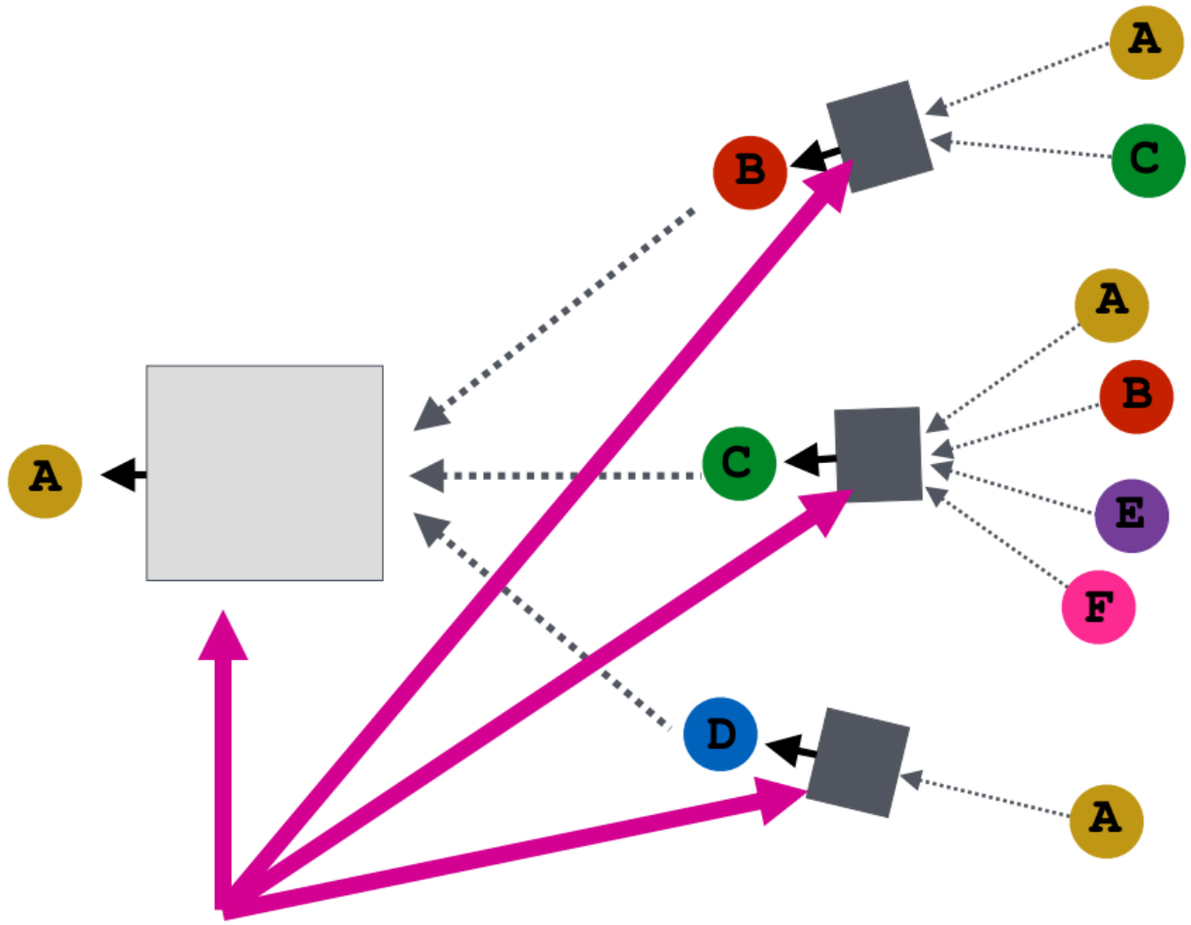
GNN



GCN



INPUT GRAPH



Neural networks

Honorable Mentions

- CapsuleNet
- Deep Belief Networks
- Neural Turing Machine
- Other Generative Model Architectures
- …

Many more …

Until next time

Questions?

Twitter: @ankurbohra

Feedback: https://tinyurl.com/pydata-feedback