Deep Learning in Wilderness

Daesoo Lee

NTNU

NTNU
Norwegian University of Science and Technology

04/2024

Overview

Overview

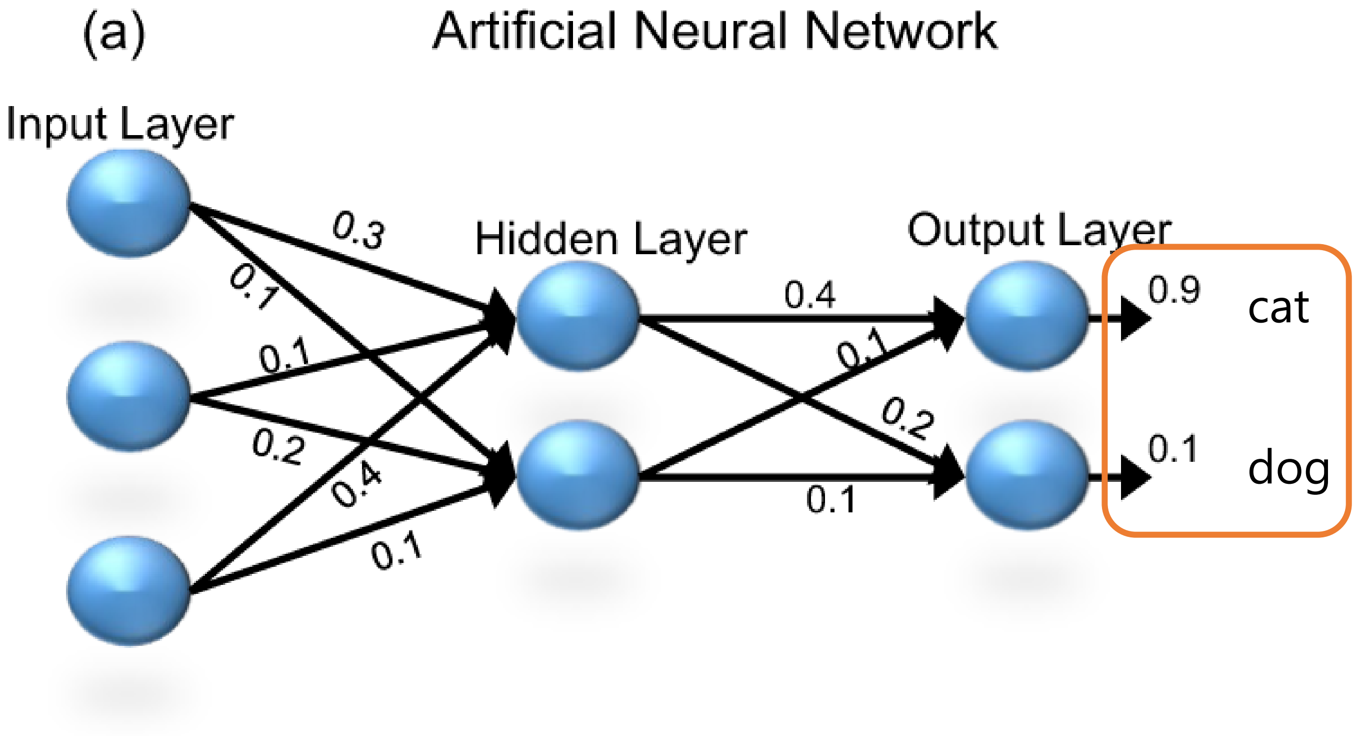
- Extension of FeedForward Network (FFN)
- Why Convolutional Neural Network (CNN) for images only?
- Goodbye Recurrent Neural Network (RNN) for sequence modeling

- Active Study Fields in Deep Learning

Extension of FeedForward Network (FFN)

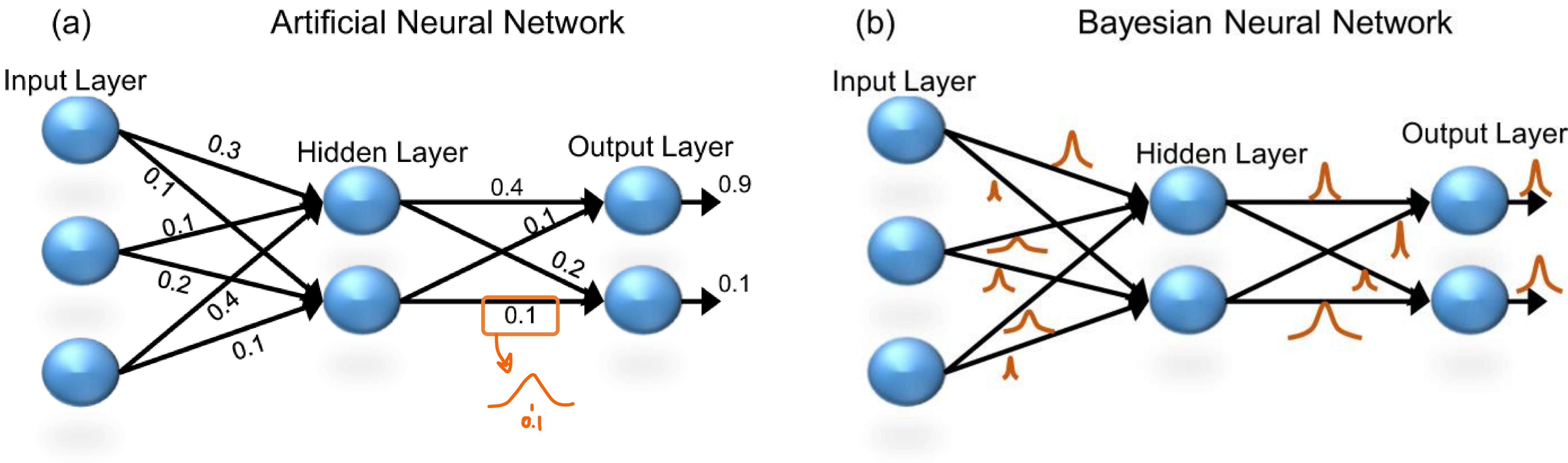
Extension of FeedForward Network (FFN)

FFN, are we confident with the prediction?

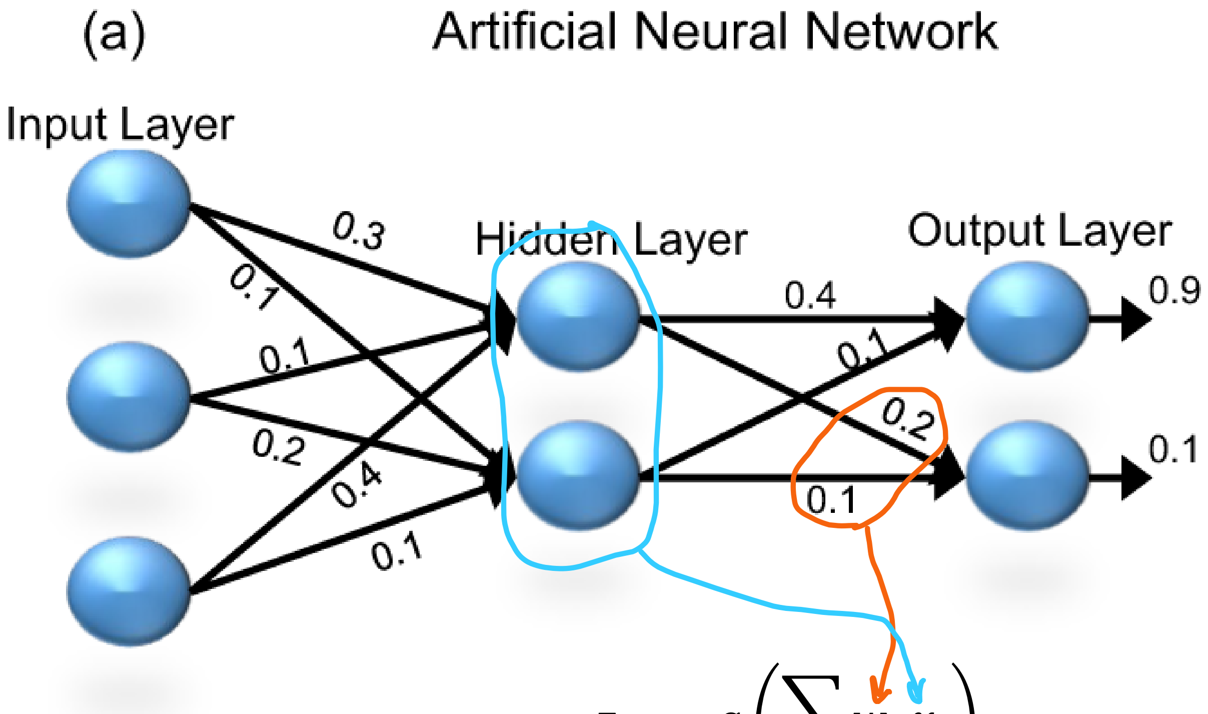


Extension of FeedForward Network (FFN)

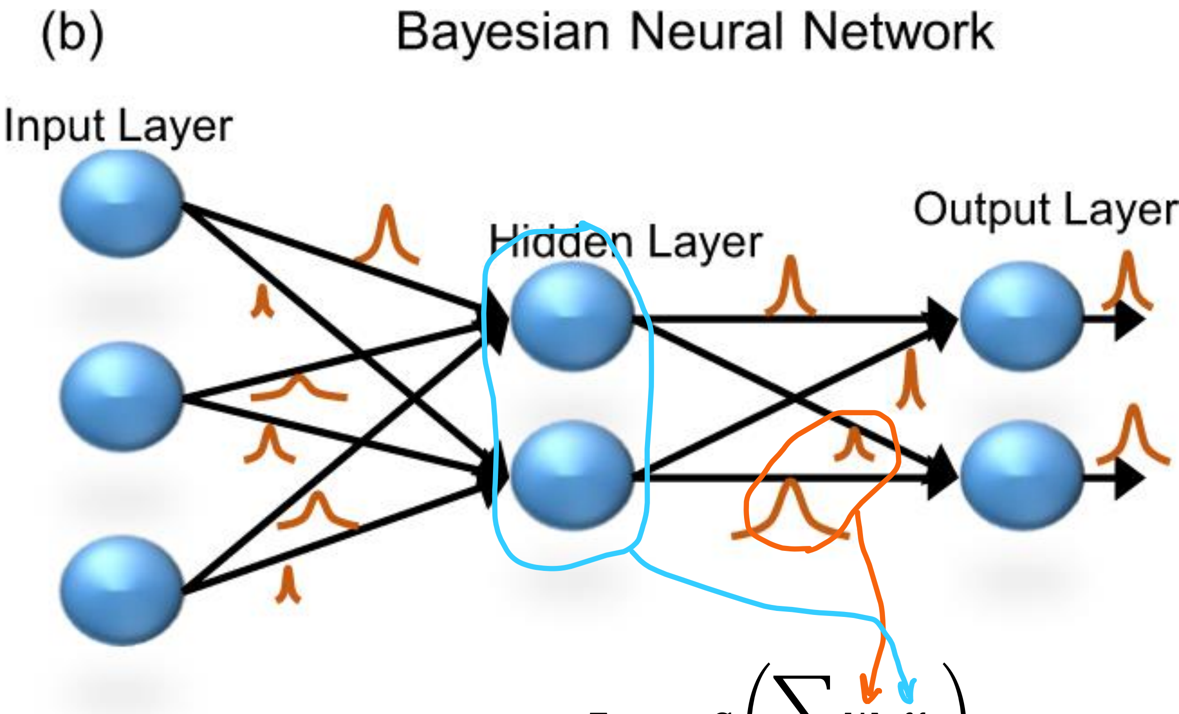
Let's admit that there's uncertainty in the model



Extension of FeedForward Network (FFN)



$$z_m = g\left(\sum_j w_j x_j\right)$$



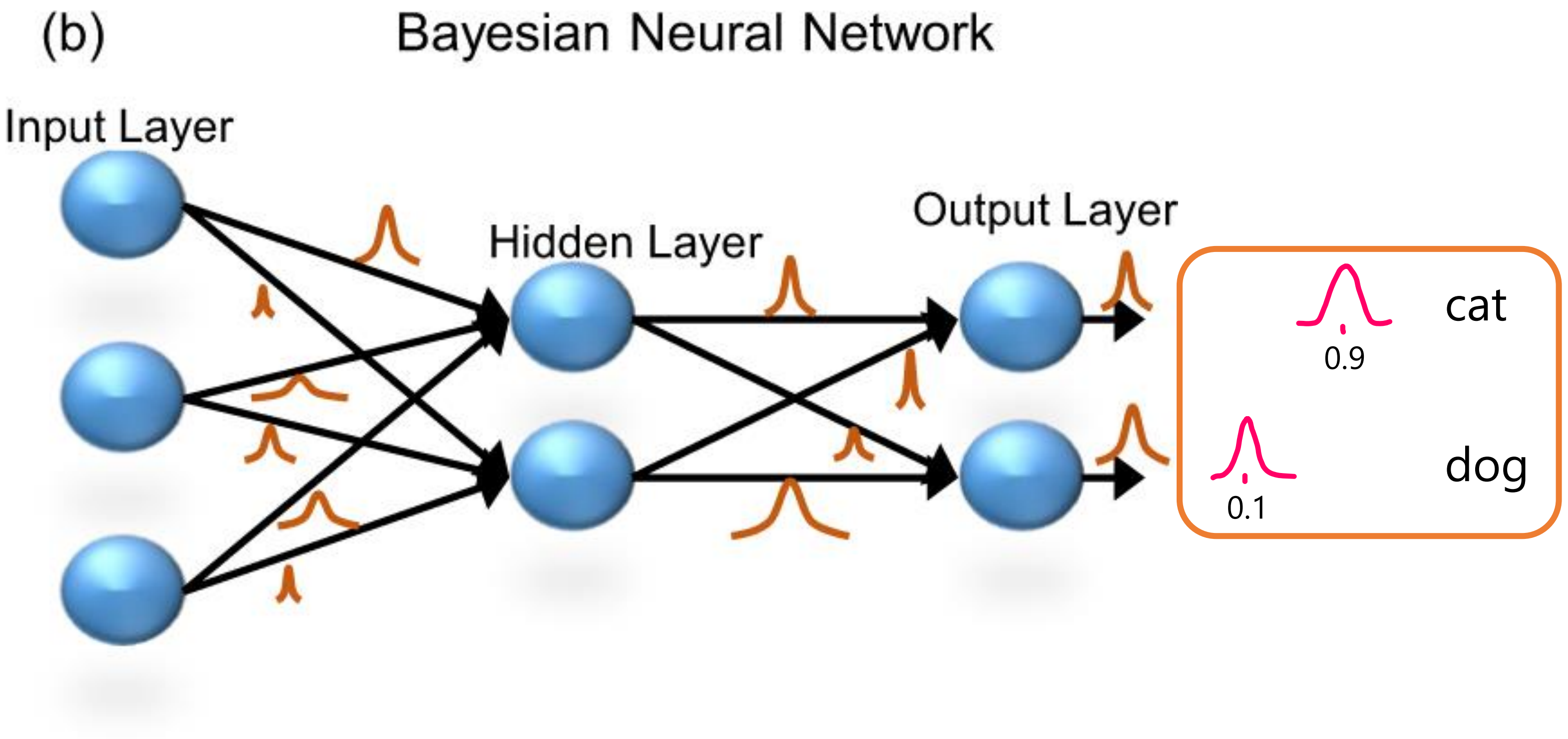
$$z_m = g\left(\sum_j w_j x_j\right)$$

$$w_j \sim p(w_j) = N(\mu, \sigma^2)$$

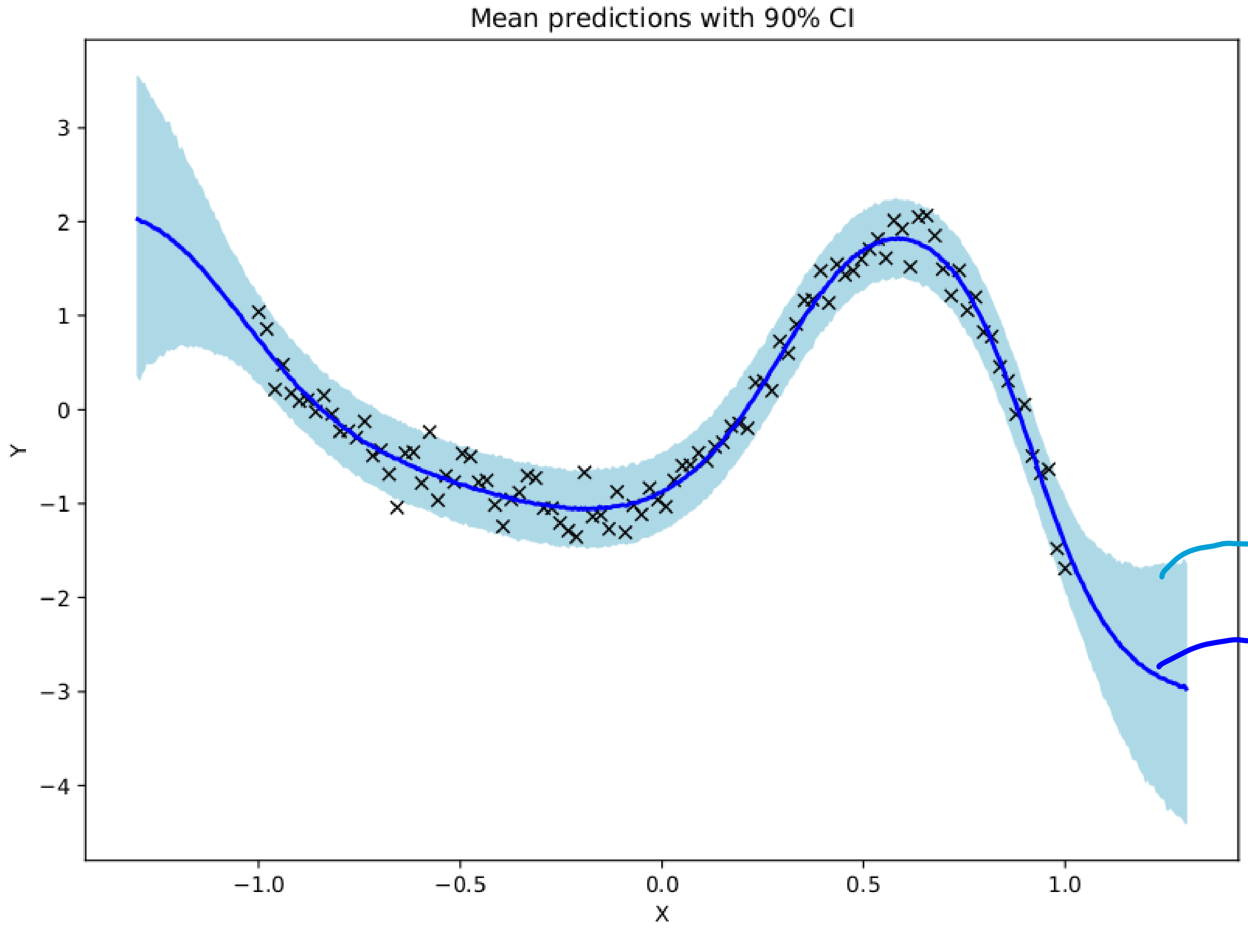
$w_j = \mu + \sigma\epsilon$ where $\epsilon \sim N(0,1)$

Extension of FeedForward Network (FFN)

Distribution of Predictions through Multiple Forward Propagations



Example Result of Bayesian Neural Net



× : samples

prediction

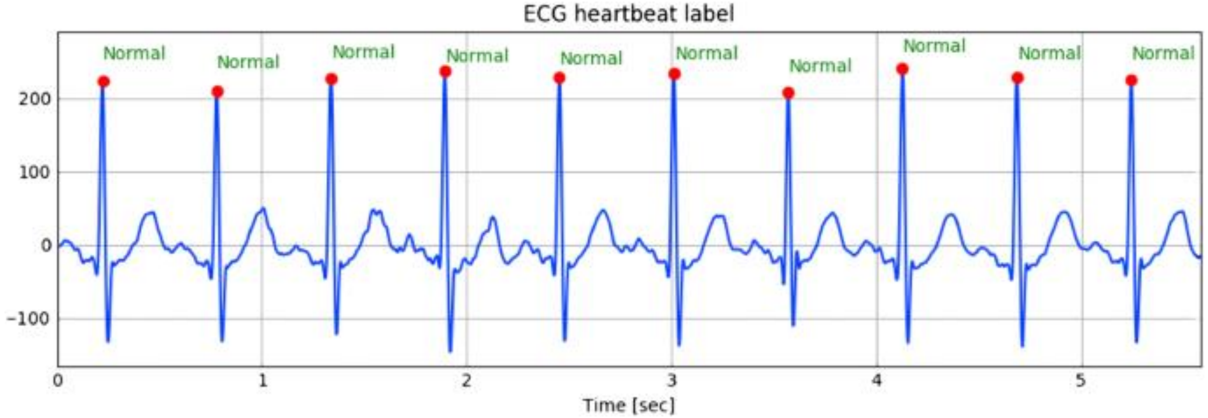
ground truth

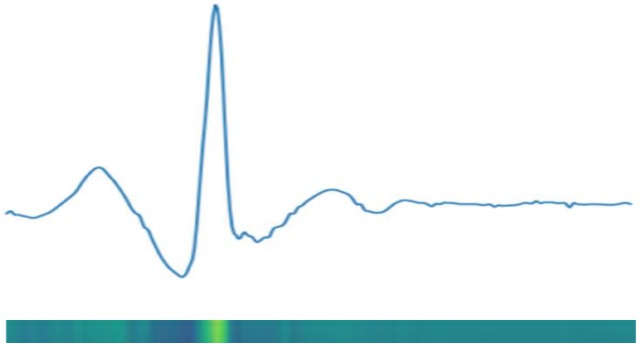
Convolutional Neural Network (CNN) for Images Only?

Convolutional Neural Network (CNN) for Images Only?

The answer is No

2D image



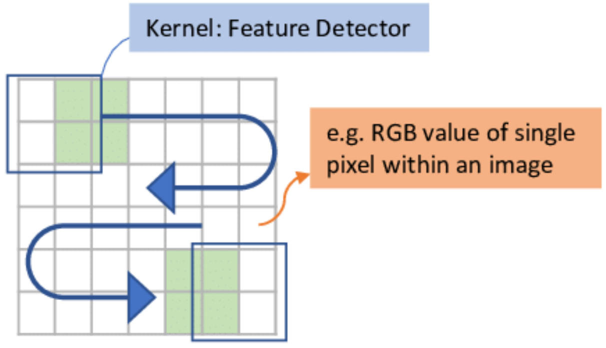


1D image

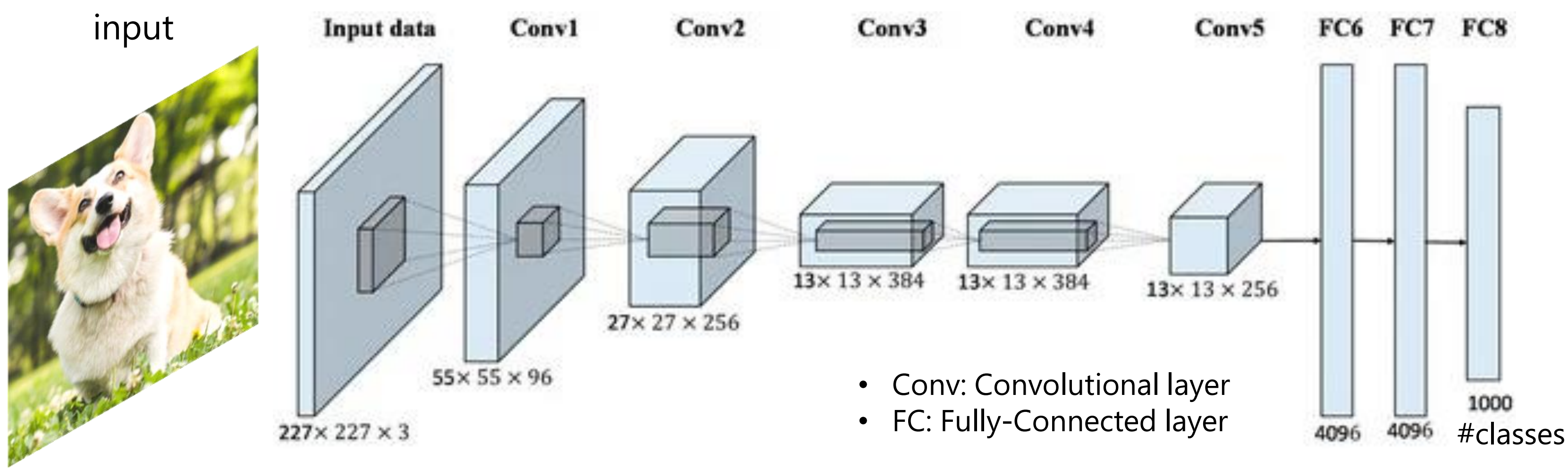
Convolutional Neural Network (CNN) for Images Only?

CNN for Time Series Processing

- 2D convolutional layer



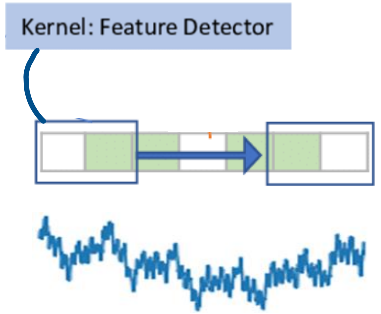
2D CNN

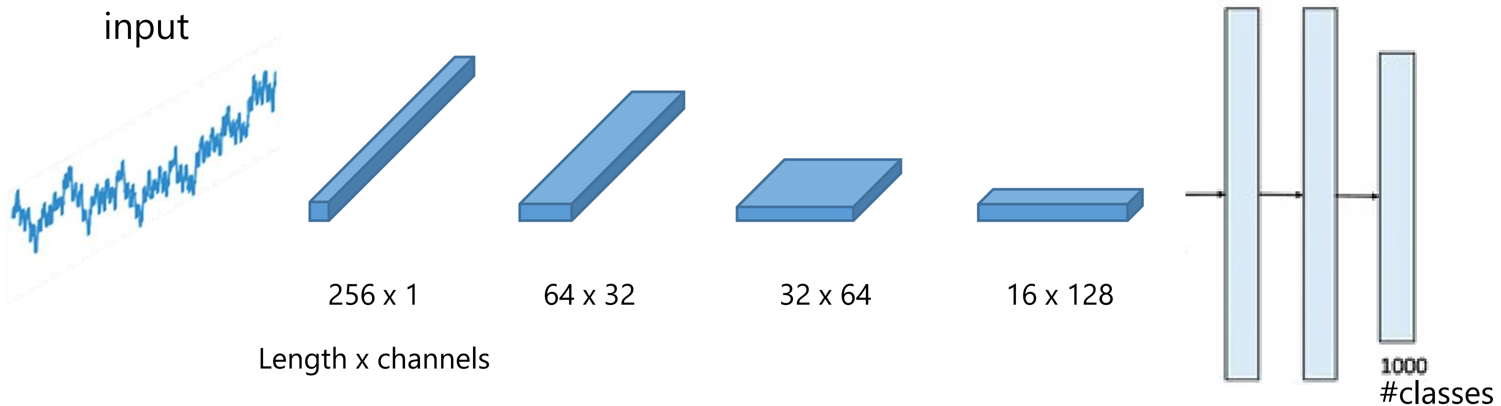


- Conv: Convolutional layer
- FC: Fully-Connected layer

Height x width x channels

- 1D convolutional layer





Length x channels

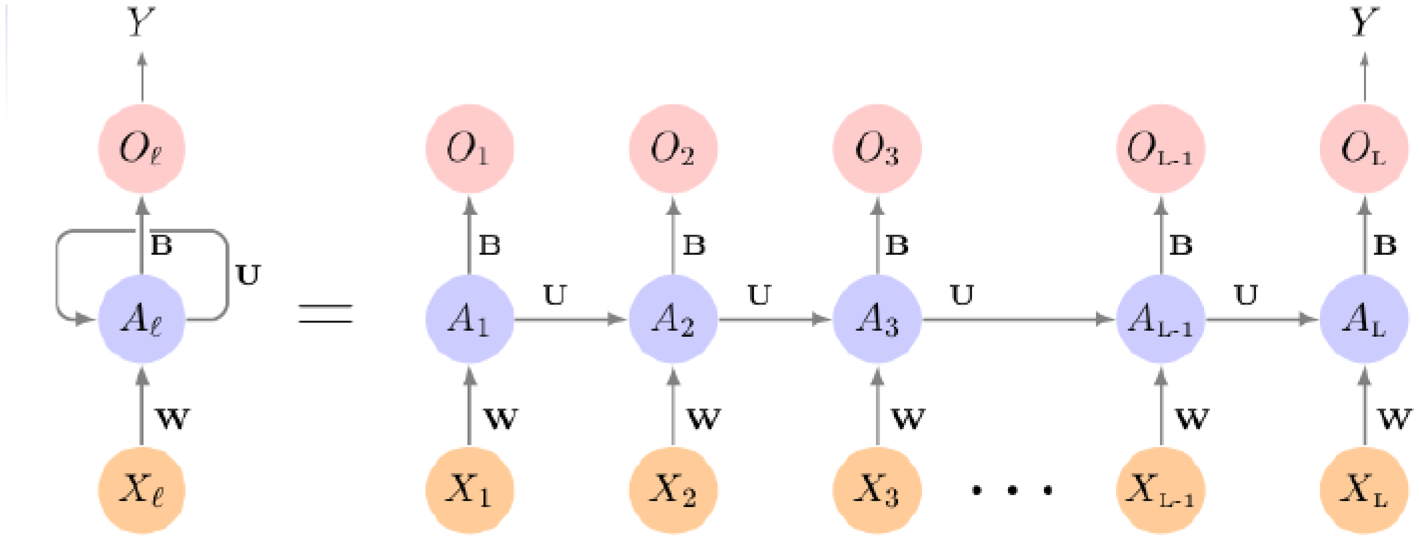
Convolutional Neural Network (CNN) for Images Only?

Popularity of using CNN for time series processing

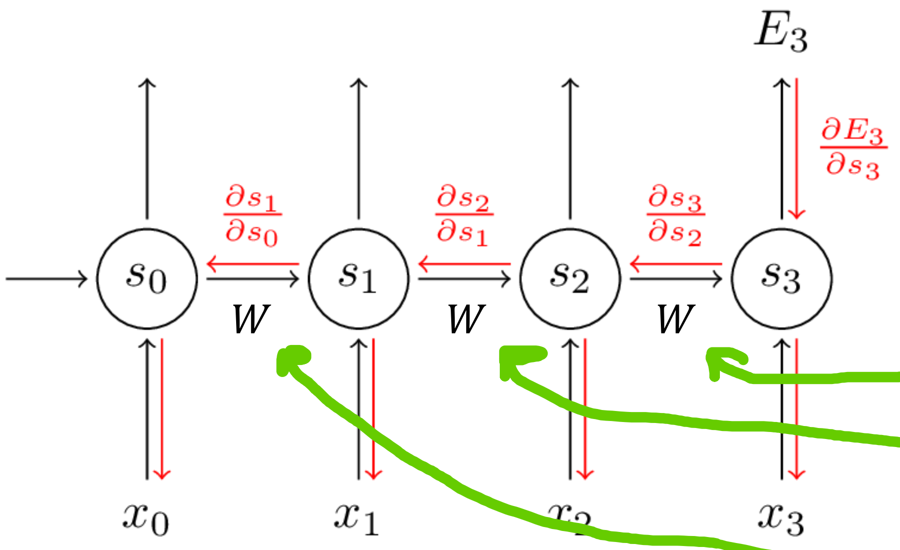
- So popular that CNN is dominantly used for this task in the literature.

Goodbye Recurrent Neural Network (RNN) for Sequence Modeling

RNN


Y
O_ℓ
B
U
A_ℓ
W
X_ℓ
=
O_1 O_2 O_3 O_{L-1} O_L
B B B B B
U U U U
A_1 A_2 A_3 A_{L-1} A_L
W W W W W
X_1 X_2 X_3 ... X_{L-1} X_L
Y

Limitations of RNN (1)



Error $E_3 = L_3(y, \hat{y})$

$$\frac{\partial E_3}{\partial W} = \frac{\partial E_3}{\partial \hat{y}} \frac{\partial \hat{y}}{\partial W}$$

$$= \frac{\partial E_3}{\partial \hat{y}} \frac{\partial \hat{y}}{\partial s_3} \frac{\partial s_3}{\partial W}$$

$$= \frac{\partial E_3}{\partial \hat{y}} \frac{\partial \hat{y}}{\partial s_3} \frac{\partial s_3}{\partial s_2} \frac{\partial s_2}{\partial W}$$

$$= \frac{\partial E_3}{\partial \hat{y}} \frac{\partial \hat{y}}{\partial s_3} \frac{\partial s_3}{\partial s_2} \frac{\partial s_2}{\partial s_1} \frac{\partial s_1}{\partial W}$$

- E : error
- s : hidden state
- W : trainable weights of RNN

https://www.analyticsvidhya.com/blog/2021/07/lets-understand-the-problems-with-recurrent-neural-networks/

Goodbye Recurrent Neural Network (RNN) for Sequence Modeling

Limitations of RNN (1)

Error $E_{1000} = L_{1000}(y, \hat{y})$

$$\frac{\partial E_{1000}}{\partial W}$$

$$= \frac{\partial E_{1000}}{\partial \hat{y}} \frac{\partial \hat{y}}{\partial s_{1000}} \frac{\partial s_{1000}}{\partial W}$$

$$+ \frac{\partial E_{1000}}{\partial \hat{y}} \frac{\partial \hat{y}}{\partial s_{1000}} \frac{\partial s_{1000}}{\partial s_{999}} \frac{\partial s_{999}}{\partial W}$$

$$+ \frac{\partial E_{1000}}{\partial \hat{y}} \frac{\partial \hat{y}}{\partial s_{1000}} \frac{\partial s_{1000}}{\partial s_{999}} \frac{\partial s_{999}}{\partial s_{998}} \frac{\partial s_{998}}{\partial W}$$

$$\dots$$

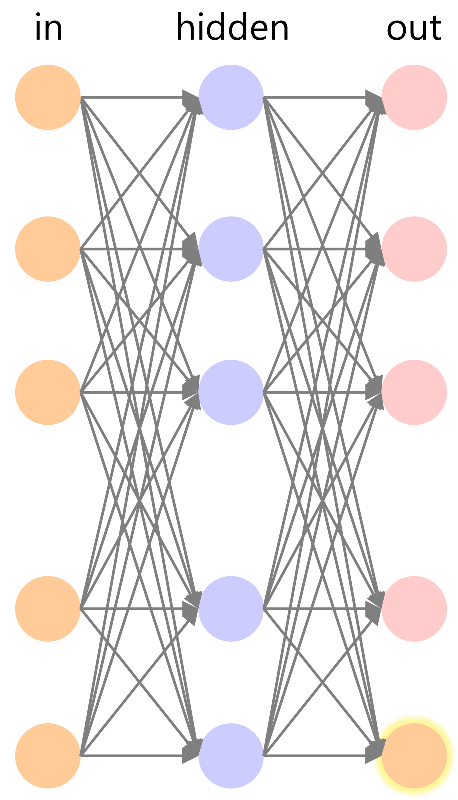
$$+ \frac{\partial E_{1000}}{\partial \hat{y}} \frac{\partial \hat{y}}{\partial s_{1000}} \frac{\partial s_{1000}}{\partial s_{999}} \frac{\partial s_{999}}{\partial s_{998}} \frac{\partial s_{998}}{\partial s_{997}} \frac{\partial s_{997}}{\partial s_{996}} \dots \frac{\partial s_2}{\partial s_1} \frac{\partial s_1}{\partial W}$$

1. *Vanishing gradient* $\left\| \frac{\partial s_i}{\partial s_{i-1}} \right\|_2 < 1$
2. *Exploding gradient* $\left\| \frac{\partial s_i}{\partial s_{i-1}} \right\|_2 > 1$

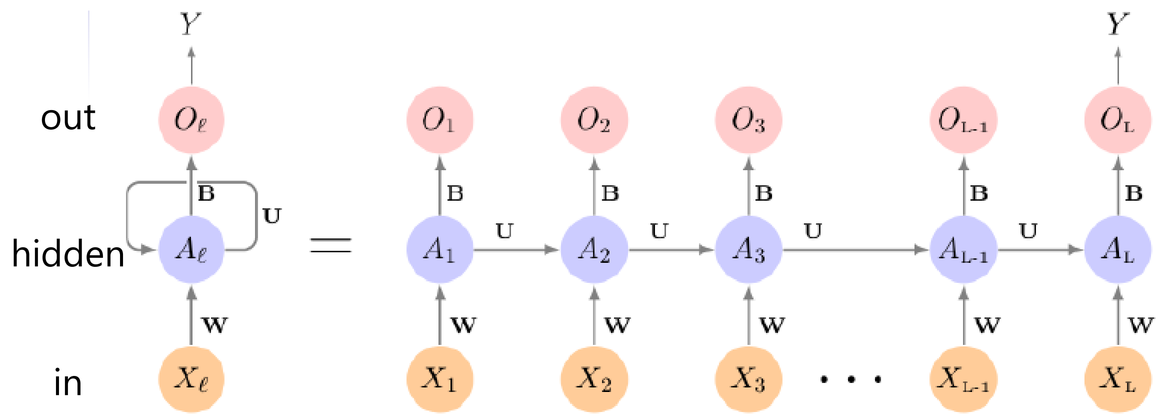
https://www.analyticsvidhya.com/blog/2021/07/lets-understand-the-problems-with-recurrent-neural-networks/

Limitations of RNN (2)

- Slow training due to the difficulty with *parallel computing*.



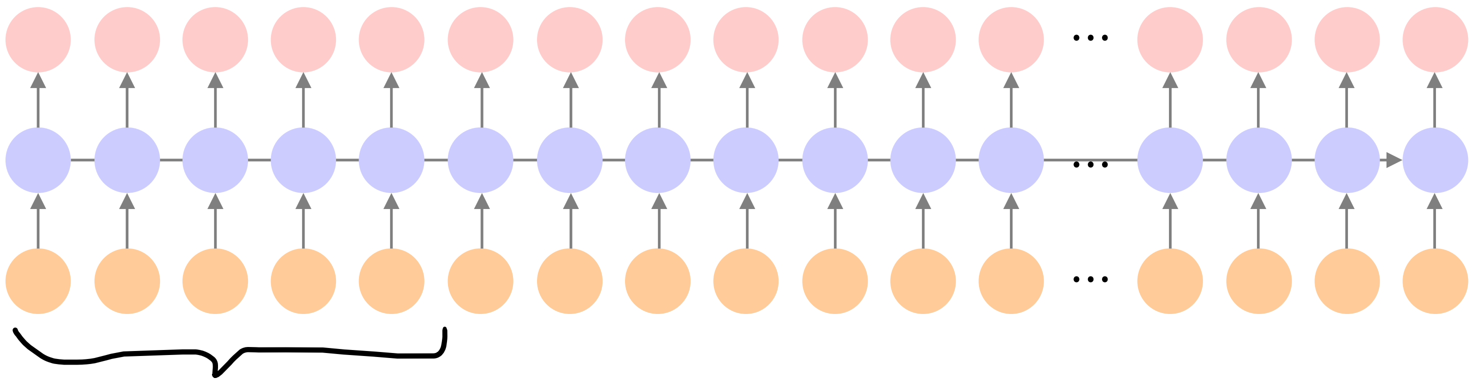
Feedforward network



RNN

Limitations of RNN (3)

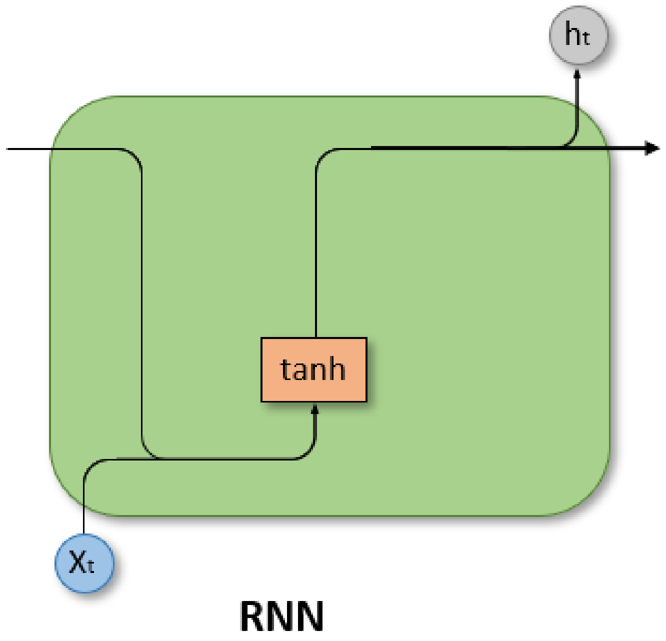
- Difficulty with processing long sequences (*i.e.*, the model forgets memories long in the past.)

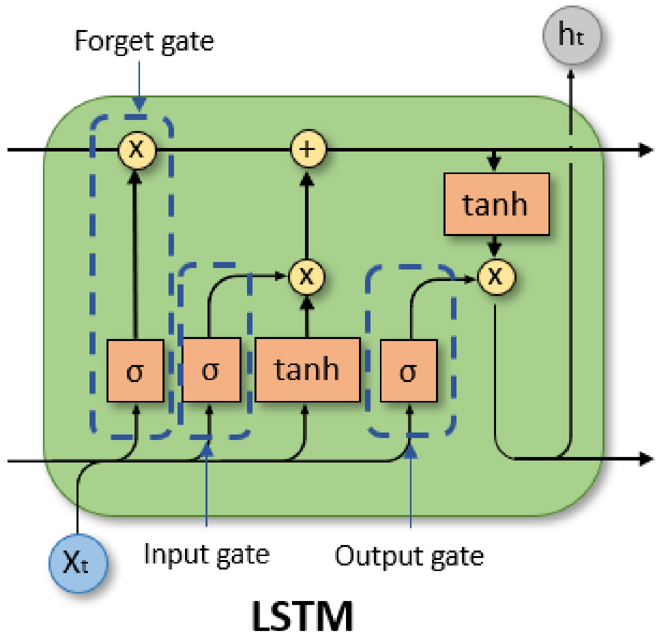


Early signals gets weak as the processing step goes on.

Long Short Term Memory (LSTM)

- To resolve the forgetting problem to some extent.





Transformer,
the replacement of RNNs

Transformer, the replacement of RNNs

A new architecture that revolutionized sequence modeling in Deep Learning

Attention Is All You Need

Ashish Vaswani* Google Brain avaswani@google.com
Noam Shazeer* Google Brain noam@google.com
Niki Parmar* Google Research nikip@google.com
Jakob Uszkoreit* Google Research usz@google.com
Llion Jones* Google Research llion@google.com
Aidan N. Gomez* † University of Toronto aidan@cs.toronto.edu
Łukasz Kaiser* Google Brain lukaszkaiser@google.com
Illia Polosukhin* ‡ illia.polosukhin@gmail.com

- introduces *Transformer*
- GPT = Transformer layers
- Transformer over RNNs

GPT: Generative Pre-trained *Transformer*.

Transformer, the replacement of RNNs

Essence of Transformer: “Attention Layer”: 1) Self-Attention

- The animal didn’t cross the street because it was too tired.

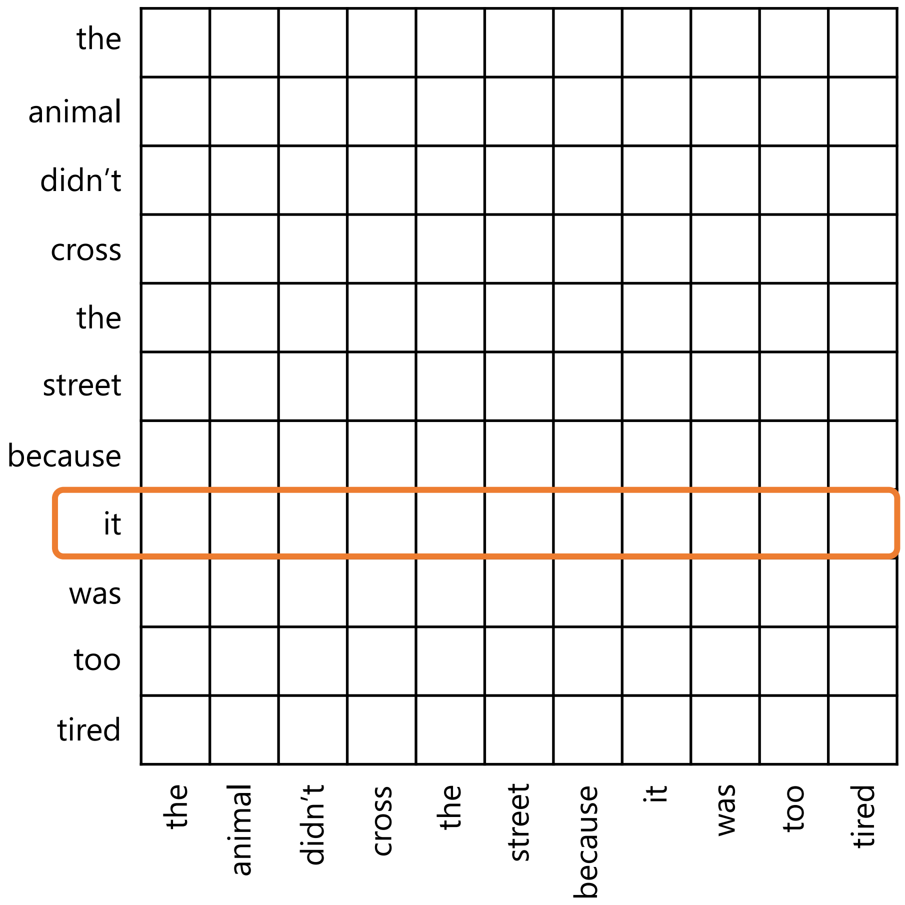
	the	animal	didn’t	cross	the	street	because	it	was	too	tired
the											
animal											
didn’t											
cross											
the											
street											
because											
it											
was											
too											
tired											

(sort of) correlation matrix between the same sentences.

Transformer, the replacement of RNNs

Essence of Transformer: “Attention Layer”: 1) Self-Attention

- The animal didn’t cross the street because it was too tired.

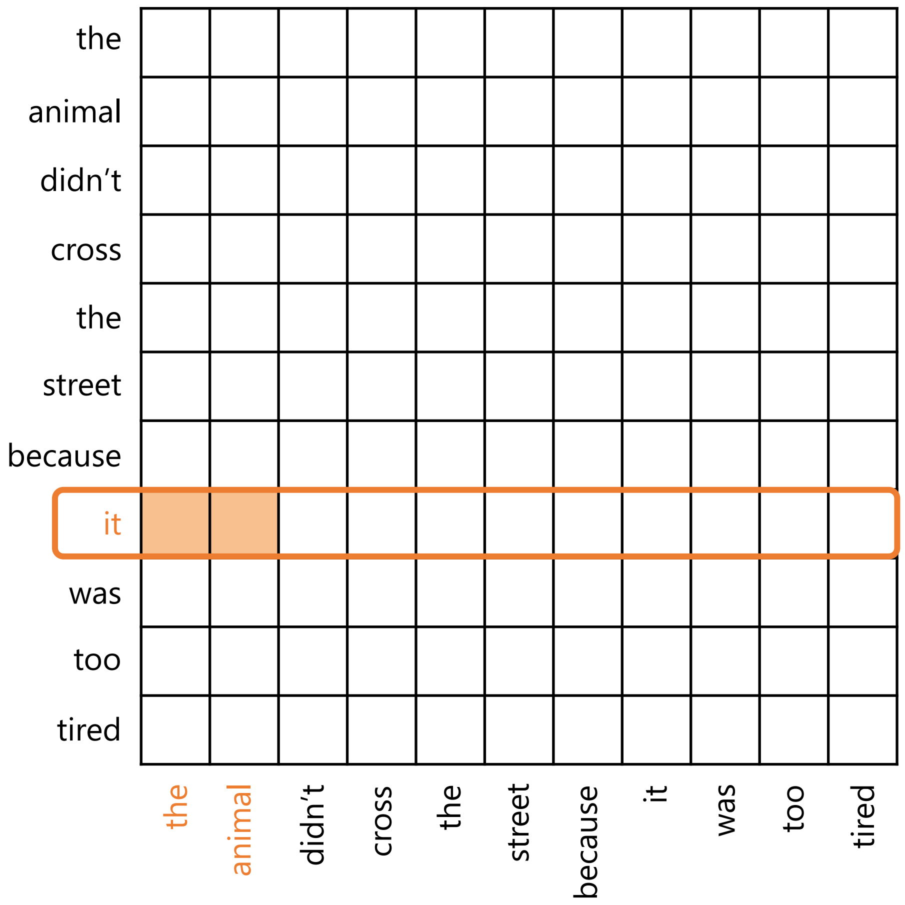


(sort of) correlation matrix between the same sentences.

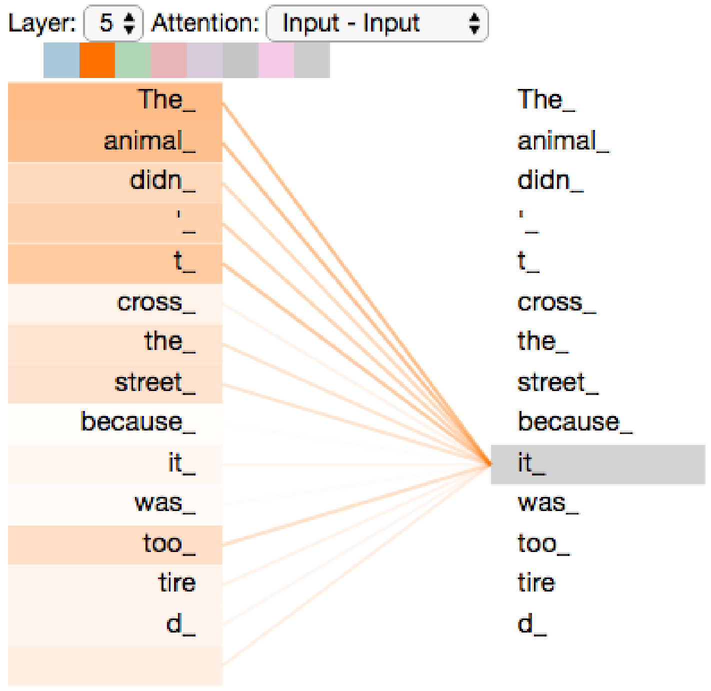
Transformer, the replacement of RNNs

Essence of Transformer: “Attention Layer”: 1) Self-Attention

- The animal didn’t cross the street because it was too tired.



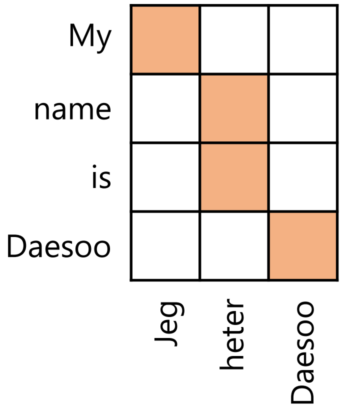
(sort of) correlation matrix between the same sentences.



Transformer, the replacement of RNNs

Essence of Transformer: “Attention Layer”: 2) Cross-Attention

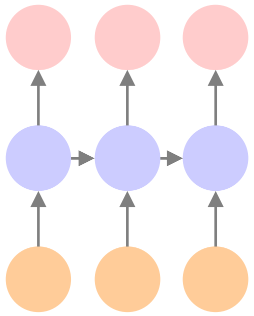
- My name is Daesoo ⬄ Jeg heter Daesoo



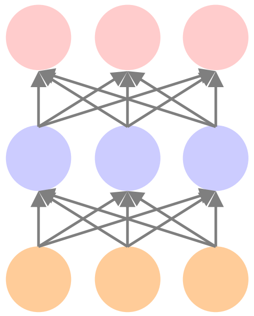
used for machine translation (e.g., google translator)

Transformer, the replacement of RNNs

RNNs vs Transformer



RNN



Transformer (block)

	RNN	Transformer
Parallel computing	x	o
Vanishing gradient	o	x
Forgetting problem	o	x

Transformer, the replacement of RNNs

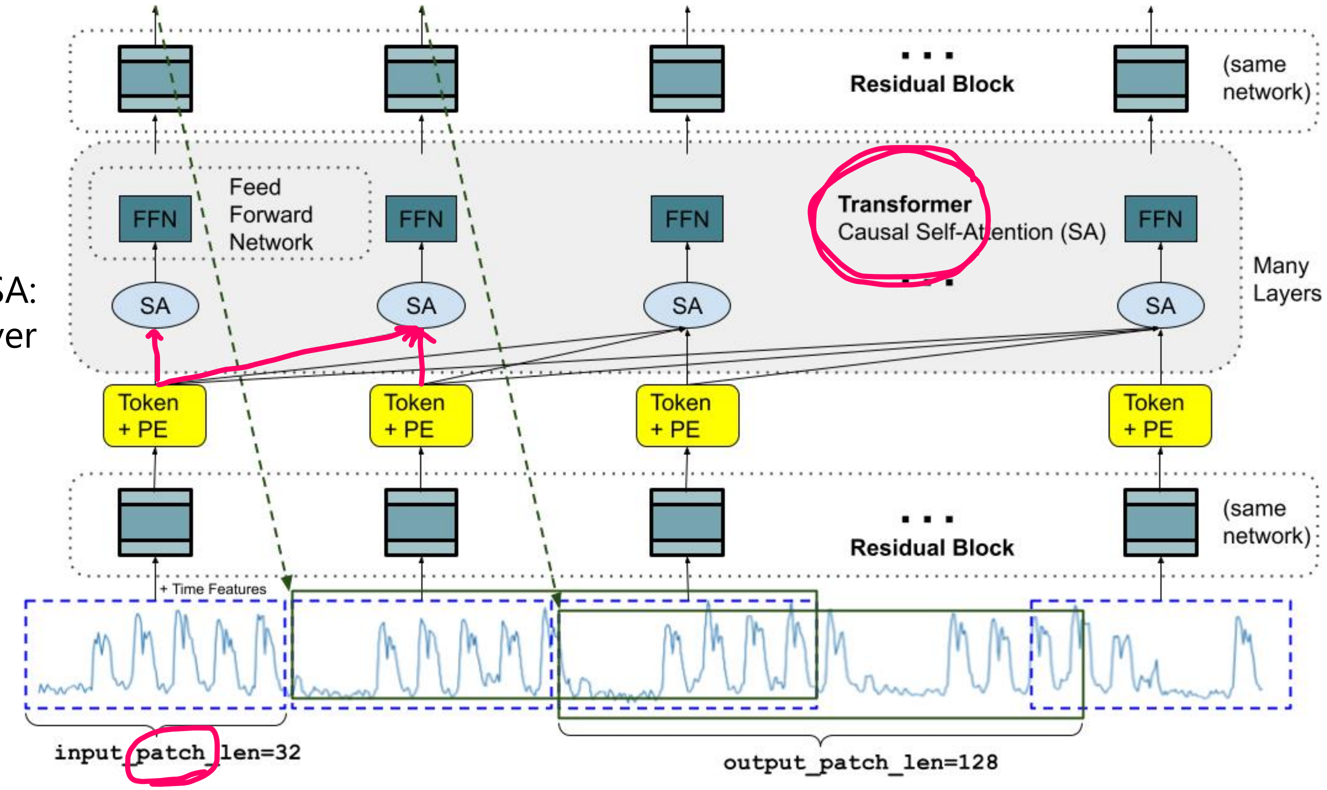
Transformer in use in real life

- Google translator
- ChatGPT
- DALL-E

Transformer, the replacement of RNNs

Transformers for Time Series Forecasting

SA:
Self-Attention Layer



Das, Abhimanyu, et al. "A decoder-only foundation model for time-series forecasting." arXiv preprint arXiv:2310.10688 (2023). Google Research

Transformer, the replacement of RNNs

Transformers for Image Modeling

treat the image as a sequence of 9.

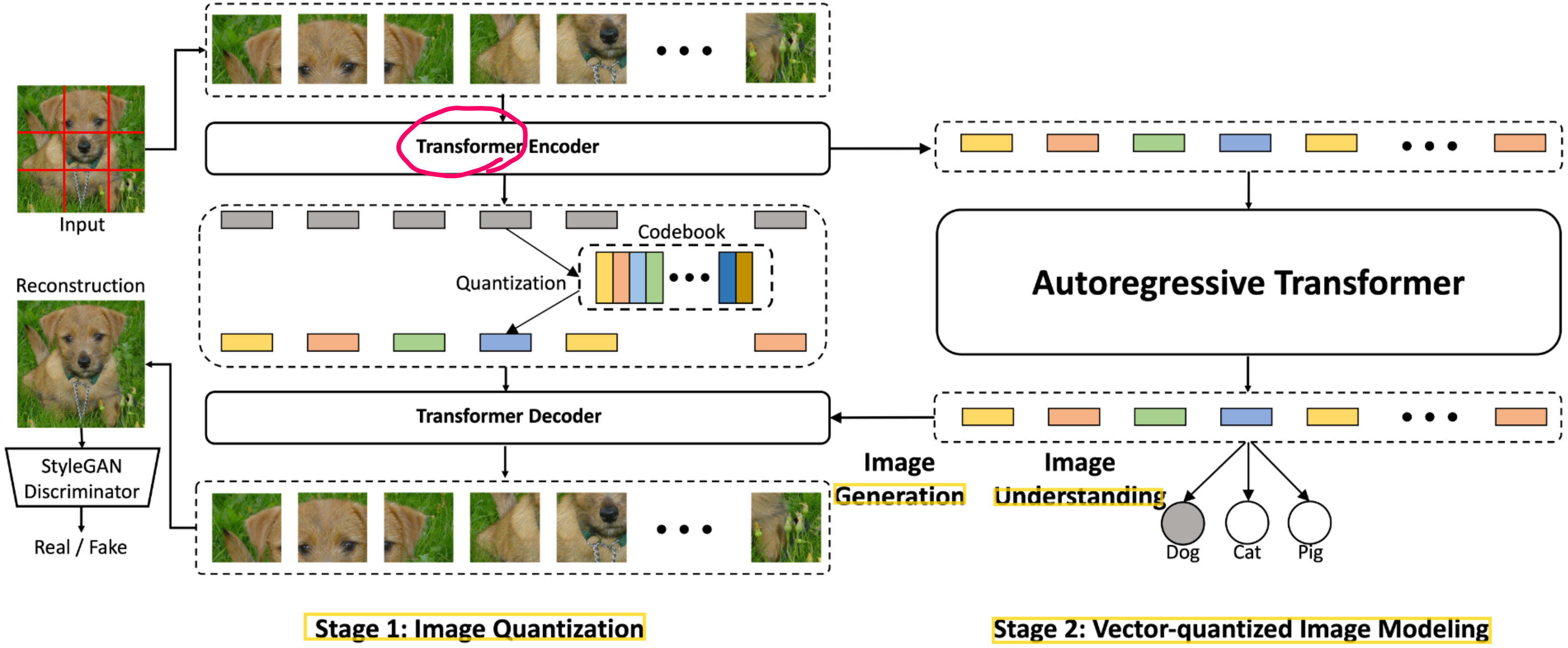


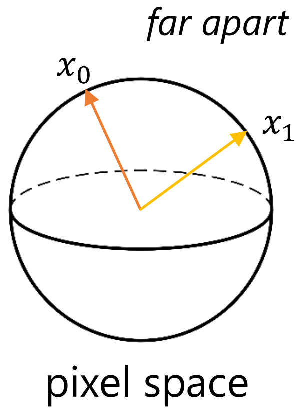
Figure 1: Overview of ViT-VQGAN (left) and Vector-quantized Image Modeling (right) for both image generation and image understanding.

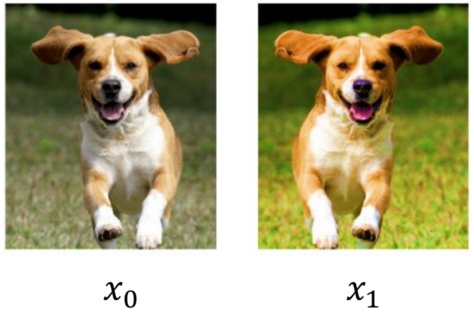
Active Study Fields in Deep Learning

Active Study Fields in Deep Learning

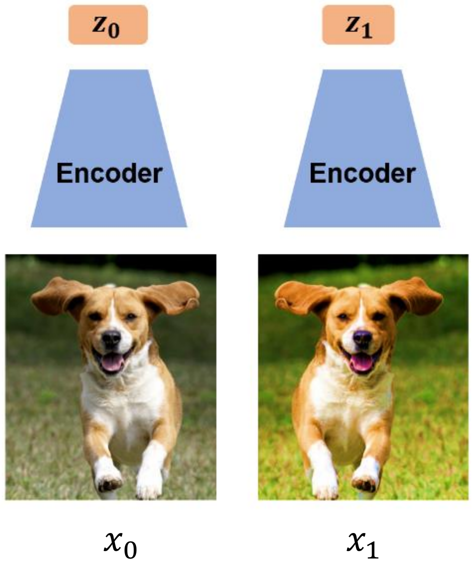
- Generative Models
- Self-supervised Learning
- Explainable AI

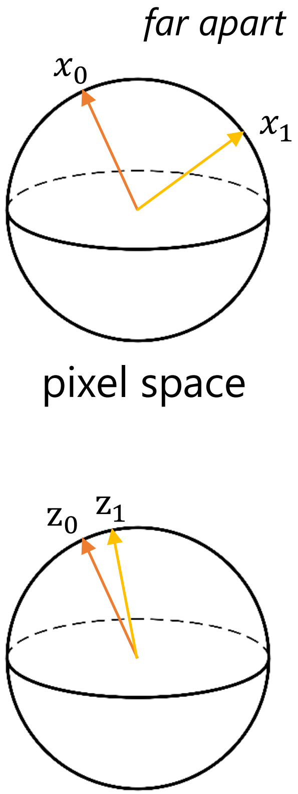
Motivation for Self-supervised Learning





Motivation for Self-supervised Learning



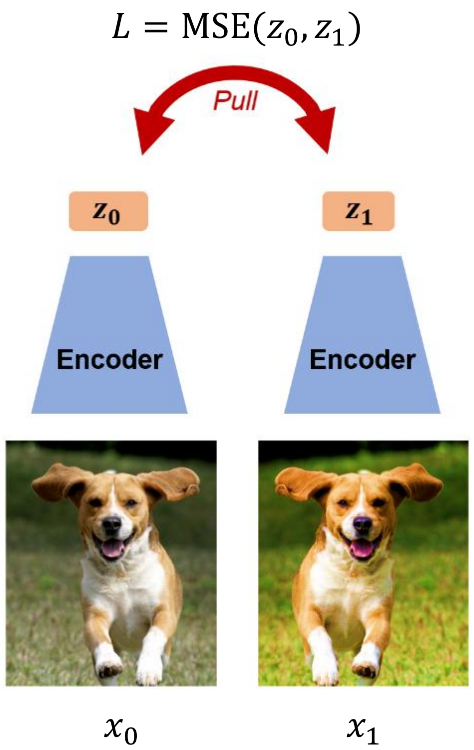


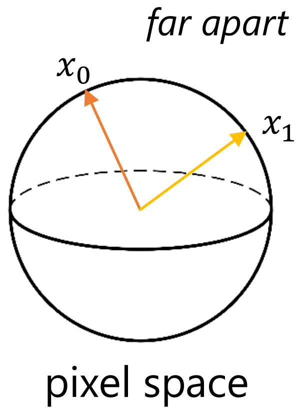
latent space $\mathcal{Z}$

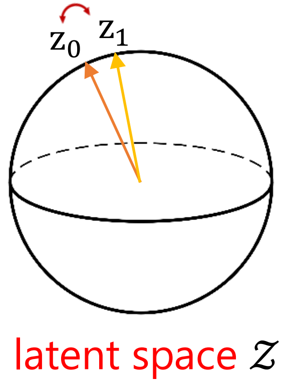
high-level semantics (*e.g.*, cat, dog, person) are captured

Active Study Fields in Deep Learning

Self-supervised Learning



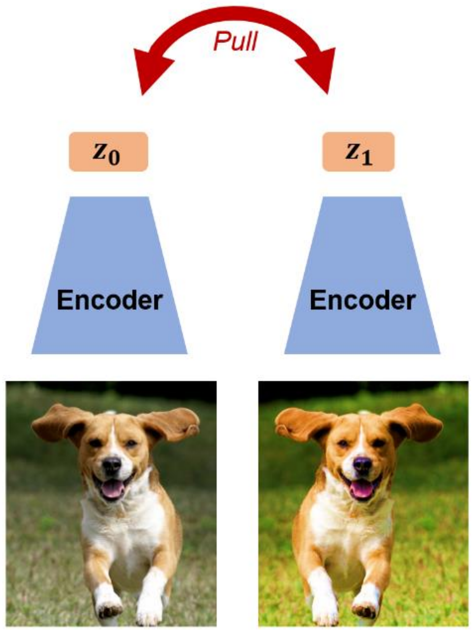




high-level semantics (*e.g.*, cat, dog, person) are captured in $\mathcal{Z}$.

Self-supervised Learning

$$L = \mathrm{MSE}(z_0, z_1)$$

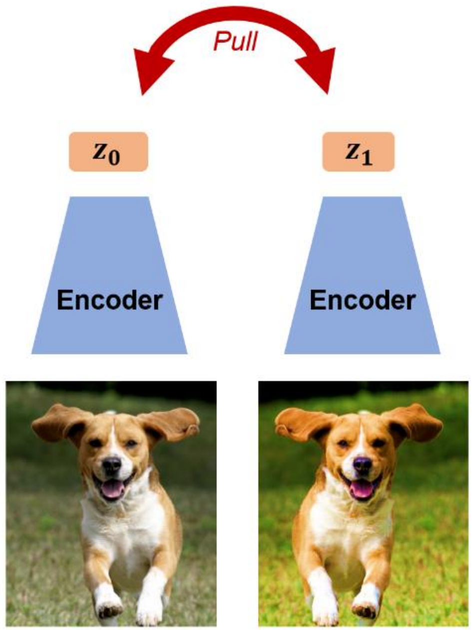


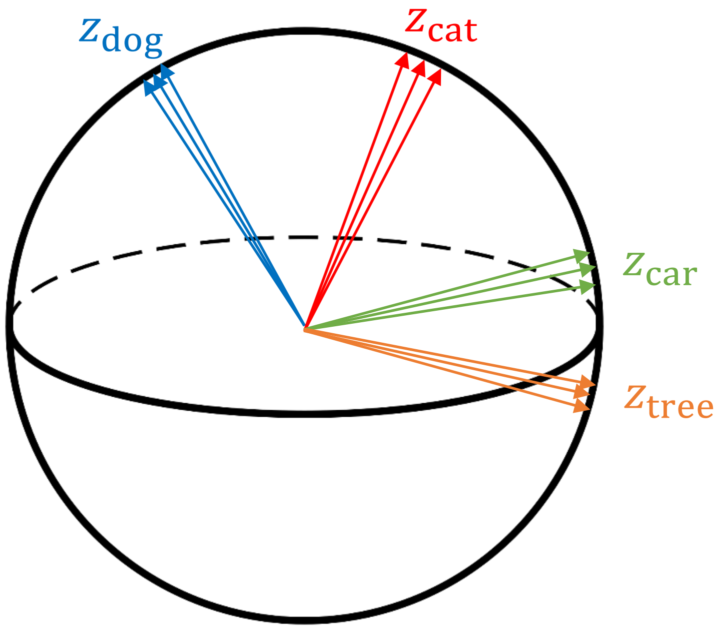
Q. any similarity to the concept of the bootstrapping in statistics?

Active Study Fields in Deep Learning

Self-supervised Learning

$$L = \text{MSE}(z_0, z_1)$$





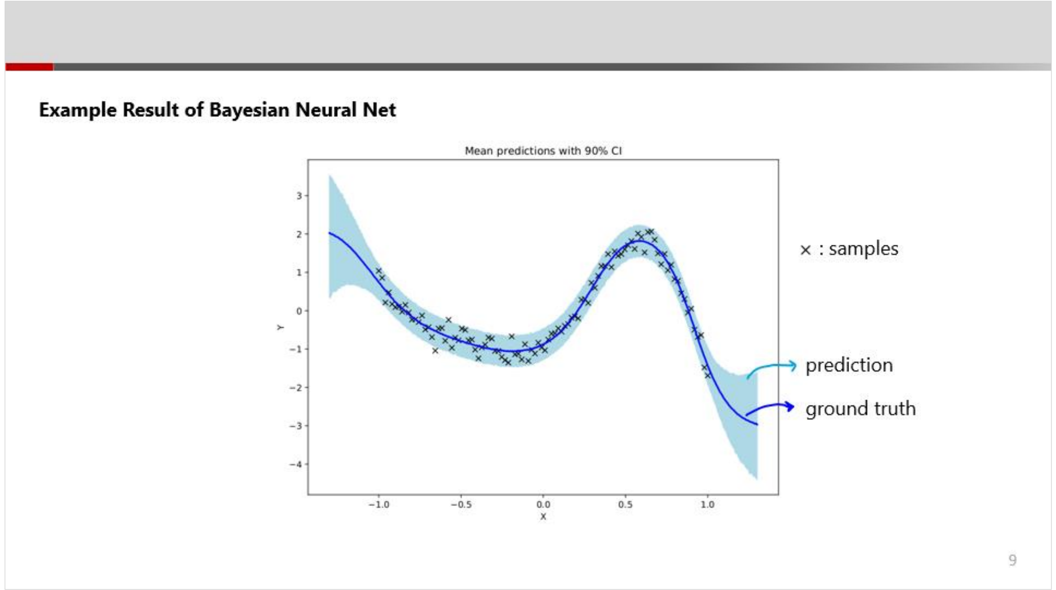
Eventually, a neural network model learns high-level visual concepts.

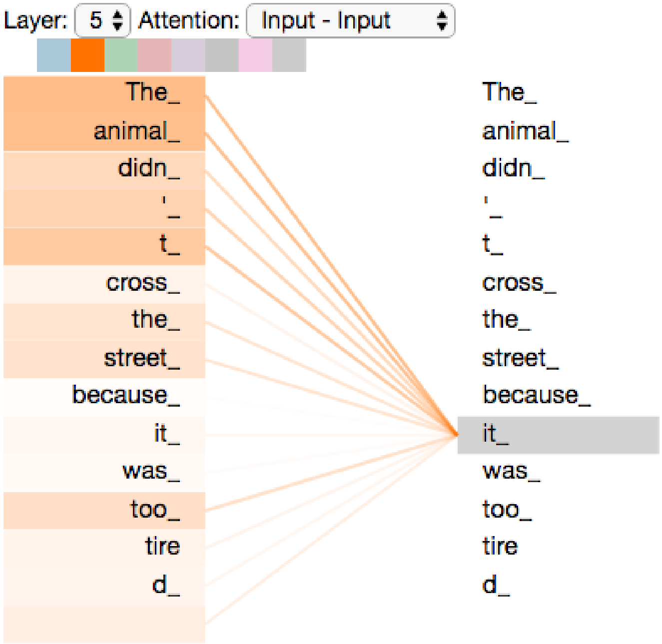
then, any task becomes easy.

Explainable AI

Explainable AI

We have already learned some of it





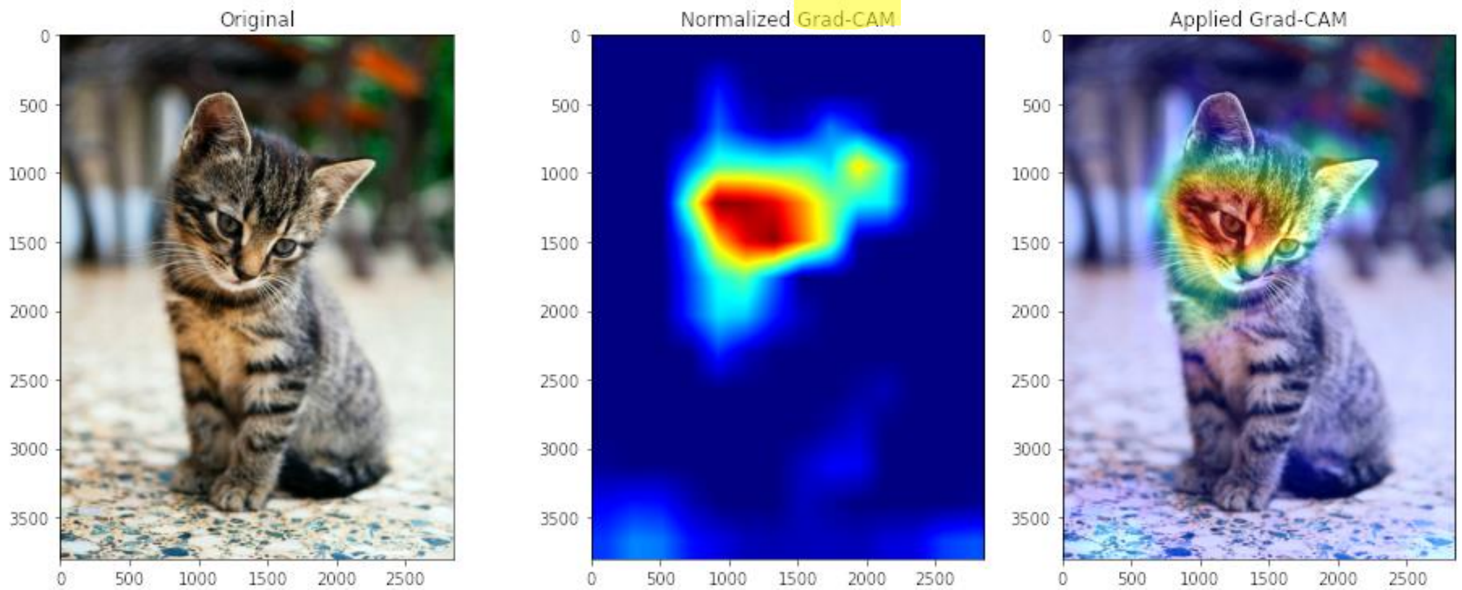
Explainable AI

Grad-CAM:
Visual Explanations from Deep Networks via Gradient-based Localization

Cited by 13,812

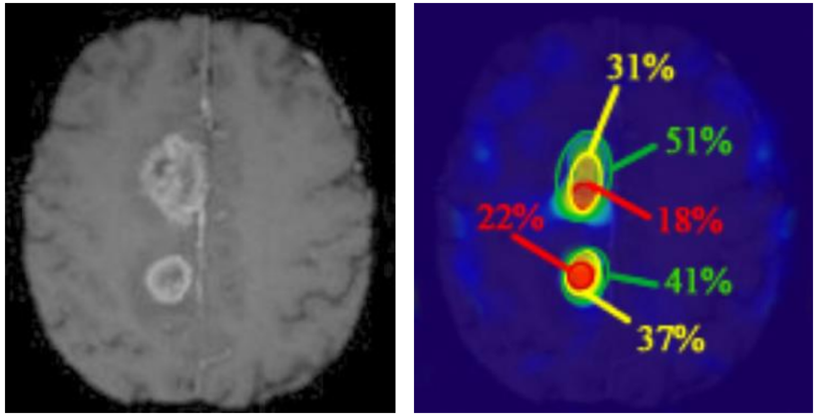
Ramprasaath R. Selvaraju[1*] Michael Cogswell[1] Abhishek Das[1] Ramakrishna Vedantam[1*]
Devi Parikh[1,2] Dhruv Batra[1,2]
[1]Georgia Institute of Technology [2]Facebook AI Research
{ramprs, cogswell, abhshkdz, vrama, parikh, dbatra}@gatech.edu

"OK, now my classification model works well.
But it'd be nicer if the model also tells me why it classified the image as that."



Explainable AI

Reasoning behind Classification



Brain tumor identification & detection

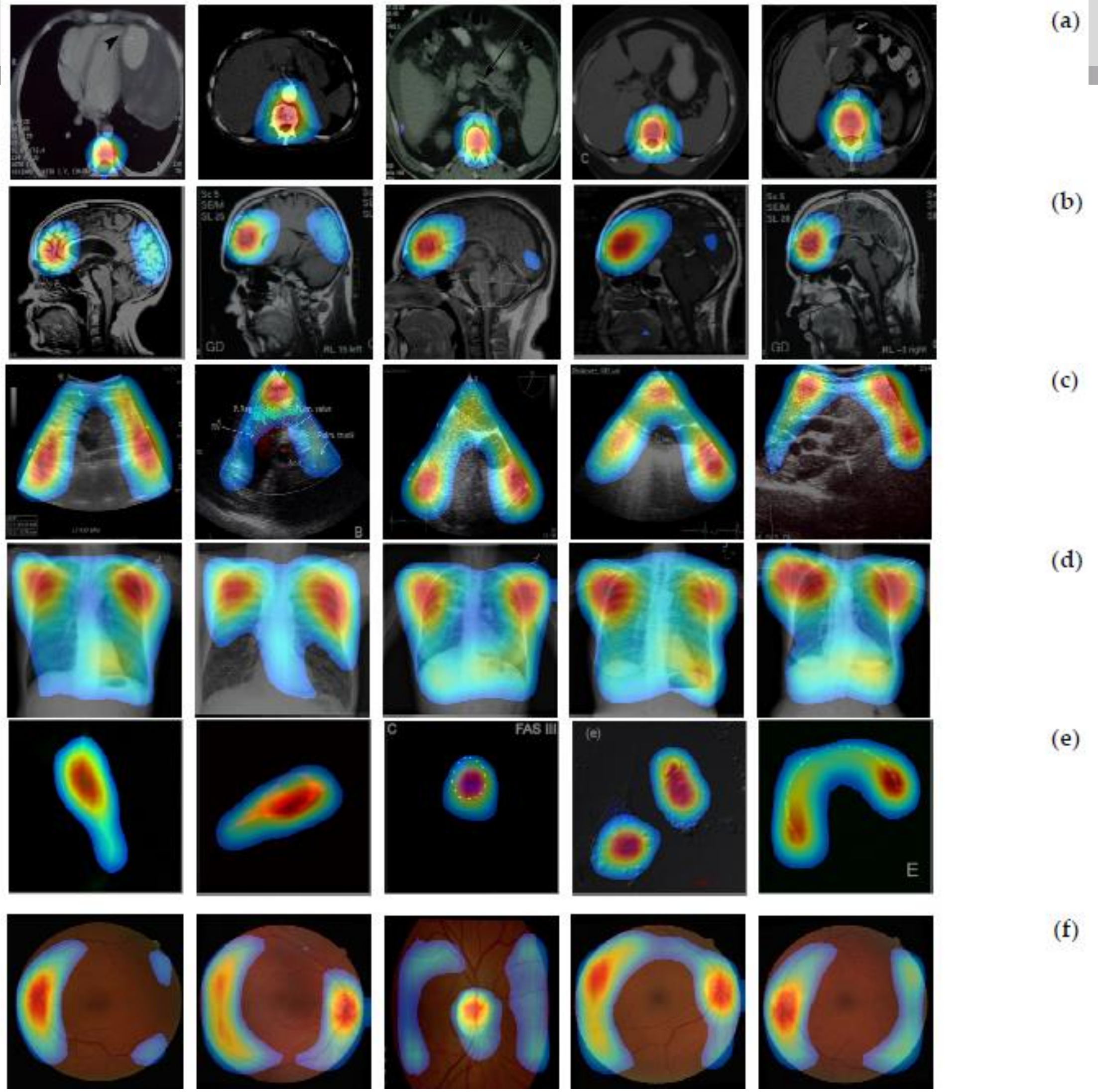
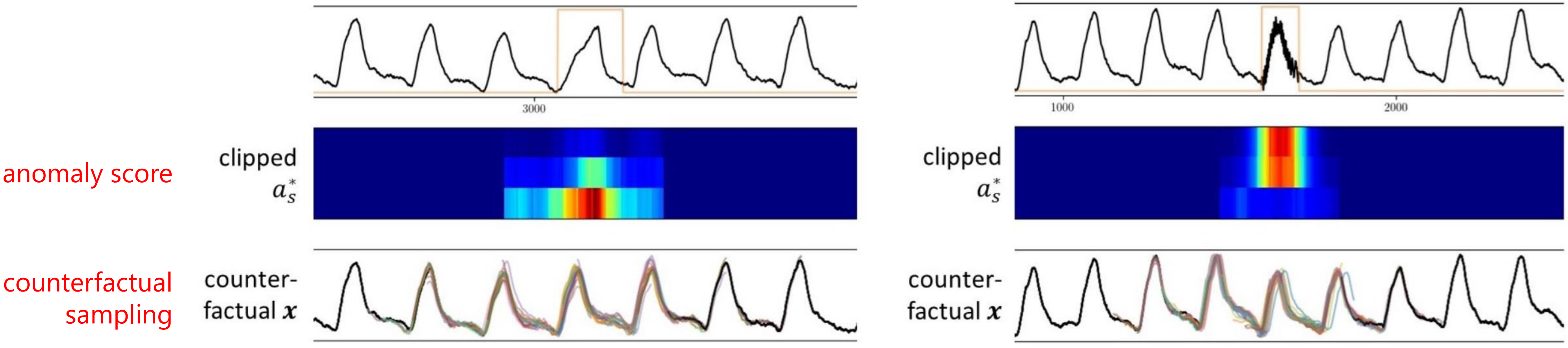


Figure 4. Heatmaps reflecting the proposed CRM for **(a)** abdomen CT, **(b)** brain MRI, **(c)** cardiac abdomen ultrasound, **(d)** chest X-ray, **(e)** fluorescence microscopy, **(f)** retinal fundoscopy, and

On the various medical images

Counterfactual Sampling in Time Series Anomaly Detection



(a) LF anomaly (b) HF anomaly

Lee, Daesoo, Sara Malacarne, and Erlend Aune. "Explainable Anomaly Detection using Masked Latent Generative Modeling." arXiv preprint arXiv:2311.12550 (2023).

Thank you!

NTNU

Norwegian University of
Science and Technology