Overview Papers

Loubna FARES

Supervisors: Abdelmalik Taleb Ahmed (IEMN DOAE) and Salim Taleb (CERAMATHS)

Polytechnic University of Hauts-de-France

March 23, 2022

Outline

Outline

Neuron

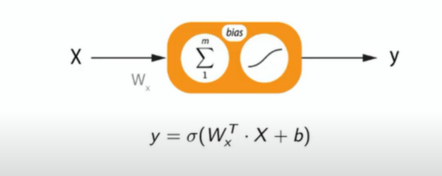


Figure: Classical neuron

Recurrent neuron

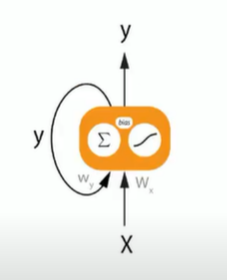
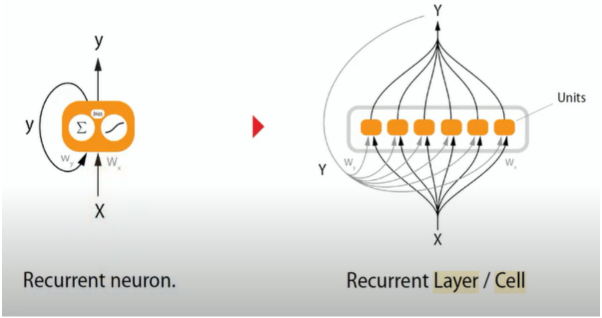


Figure: Classical neuron

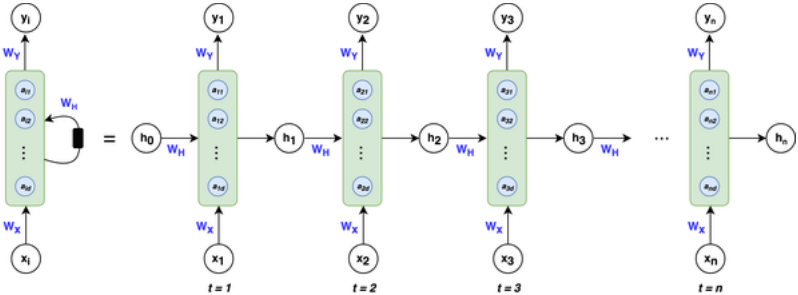
$$y_{(t)} = \sigma(W_x^T.X_{(t)} + w_y.y_{(t-1)} + b) \tag{1}$$

Recurrent Layer



$$Y_{(t)} = \phi(W_x^T.X_{(t)} + W_y^T.Y_{(t-1)} + b) \tag{2}$$

Simple RNN limitations



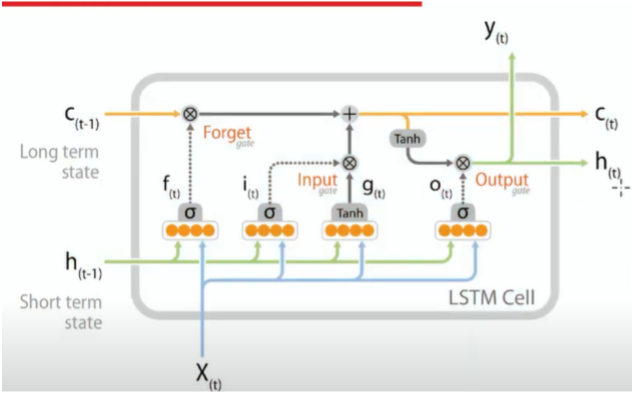
- Slow convergence
- Short memory.
- Vanishing/exploding gradient.

Outline

Long-Short-Term-Memory

LSTM Cell is composed of three big operations :

- Forget Gate
- Input Gate
- Output Gate



Long-Short-Term-Memory

- Forget Gate

$$f_t = \sigma(W_f.[h_{t-1}, x_t] + b_b) \tag{3}$$

- Input Gate

$$i_t = \sigma(W_i.[h_{t-1}, x_t] + b_i) \tag{4}$$

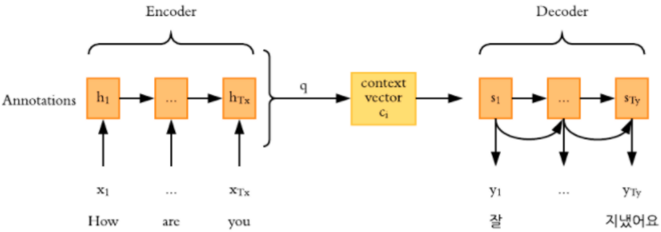
$$g_{(t)} = tanh(W_C.[h_{t-1}, x_t] + b_C) \tag{5}$$

- Output Gate

$$o_t = \sigma(W_o.[h_{t-1}, x_t] + b_o) \tag{6}$$

$$h_{(t)} = o_t * tanh(C_t) \tag{7}$$

Encoder-Decoder Architecture



Outline

Attention is All we need

- it was firstly introduced in 2017

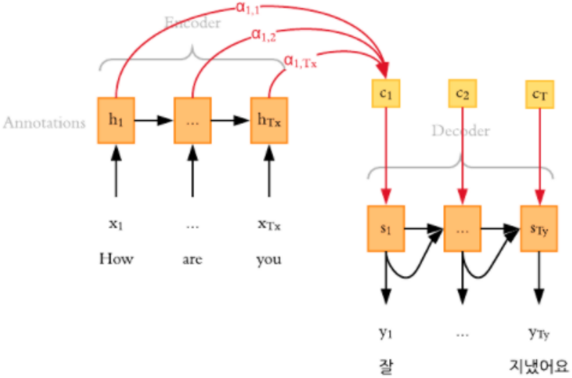
Attention Is All You Need

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

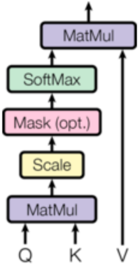
Abstract

The dominant sequence transduction models are based on complex recurrent or convolutional neural networks that include an encoder and a decoder. The best performing models also connect the encoder and decoder through an attention mechanism. We propose a new simple network architecture, the Transformer, based solely on attention mechanisms, dispensing with recurrence and convolutions entirely. Experiments on two machine translation tasks show these models to be superior in quality while being more parallelizable and requiring significantly less time to train. Our model achieves 28.4 BLEU on the WMT 2014 English-to-German translation task, improving over the existing best results, including ensembles, by over 2 BLEU. On the WMT 2014 English-to-French translation task, our model establishes a new single-model state-of-the-art BLEU score of 41.8 after training for 3.5 days on eight GPUs, a small fraction of the training costs of the best models from the literature. We show that the Transformer generalizes well to other tasks by applying it successfully to English constituency parsing both with large and limited training data.

Encoder-Decoder With Attention



Scaled Dot-Product Attention



$$\text{Attention}(Q, K, V) = \text{softmax}(\frac{QK^T}{\sqrt{d_k}})V$$

Outline

Transfomers and the mechanism of Attention

History of Transformers:

- it was firstly introduced in 2017
- in 2020 they fully replaced Convolutional architecture in image processing
- in 2021 Transformer for video understanding
- Now we are in computer Vision Revolution

Why replacing CNN with Transformers in computer vision?

- CNN are localised
- Lack of spatial information
- NLP

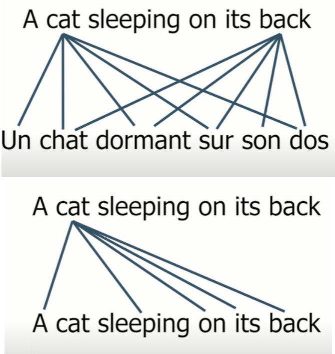
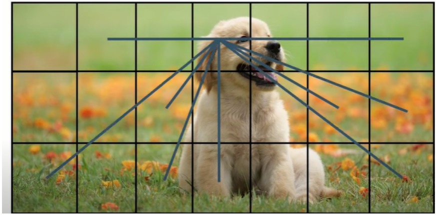
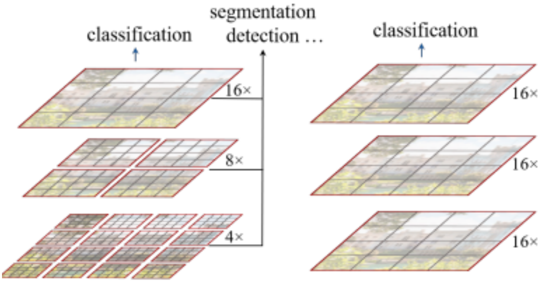


Figure: Self-Attention

Computer Vision



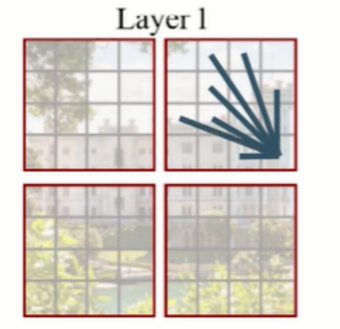
Swin Transformer: Hierarchical Vision Transformer using Shifted Windows



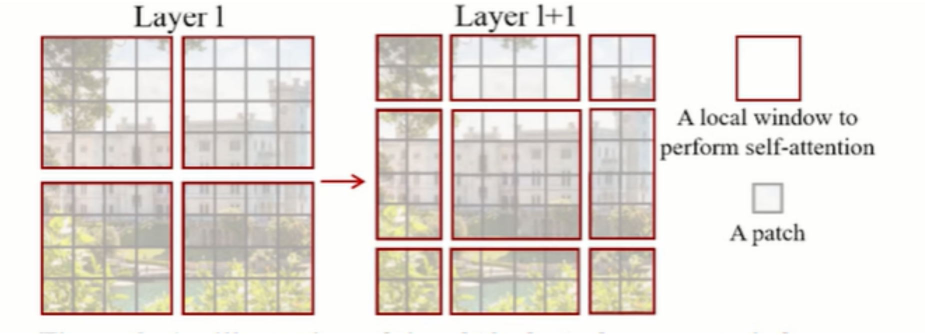
(a) Swin Transformer (ours)

(b) ViT

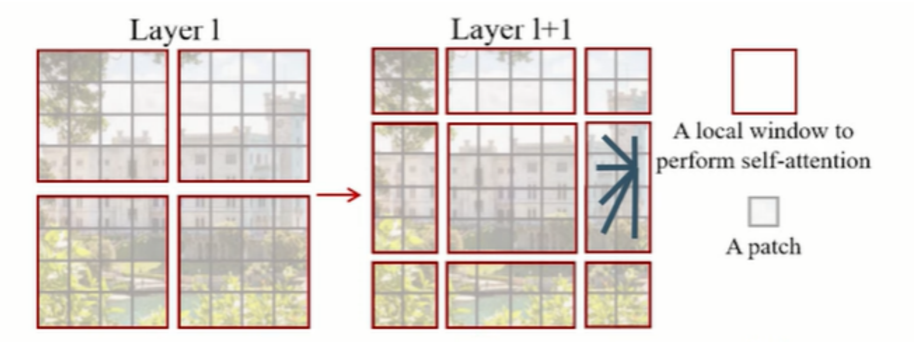
Swin Transformer

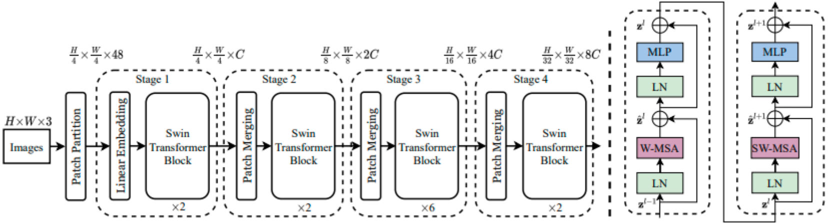


Attention


Layer l
Layer l+1
A local window to
perform self-attention
A patch

Self-Attention





(a) Architecture

(b) Two Successive Swin Transformer Blocks

Outline

References

https://www.youtube.com/watch?v=64Yl3K-4FzM&t=1295s&ab_channel=CNRS-FormationFIDLE

https://www.youtube.com/watch?v=LALTmQhVkfU&t=1557s&ab_channel=ThibaultNeveu

https://www.youtube.com/channel/UCUzGQrN-lyyc0BWTYoJM_Sg