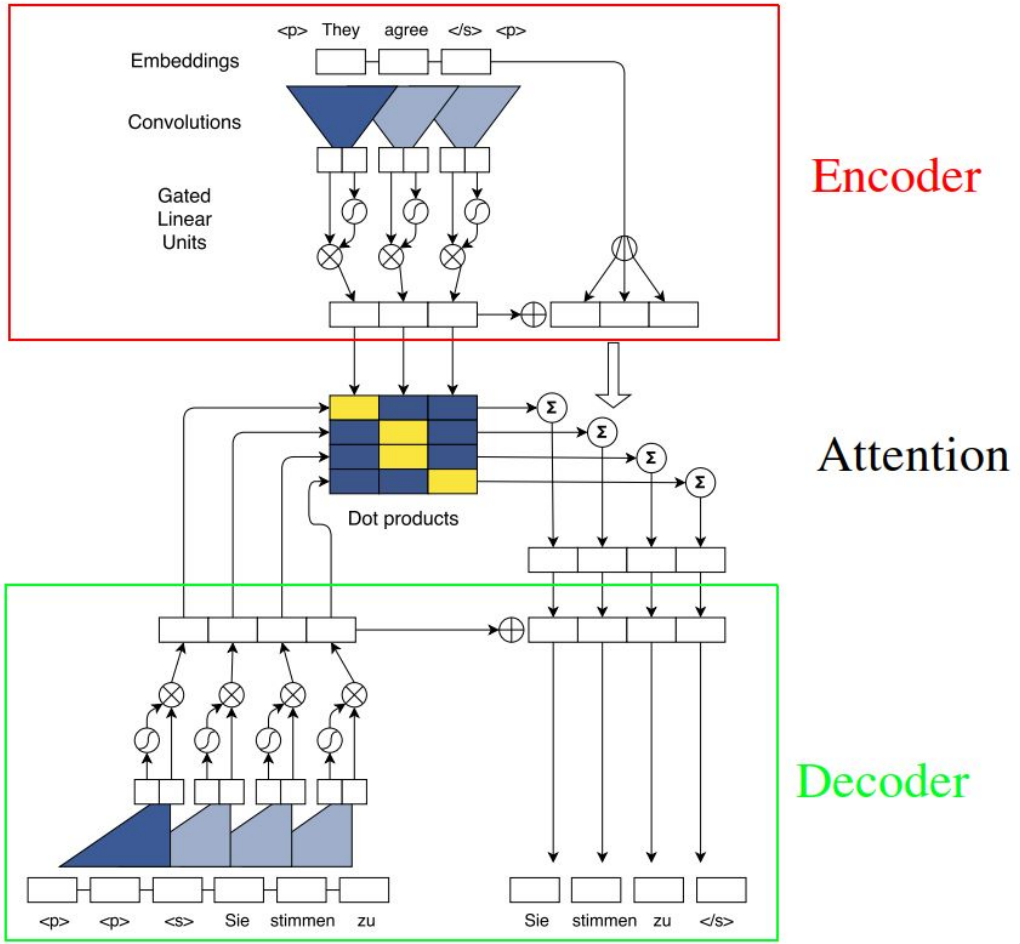
Convolutional Sequence to Sequence Learning (ConvS2S)

Presented by: **Mohamed Anwar**

ConvS2S


<p> They agree </s> <p>
Embeddings
Convolutions
Gated Linear Units
Encoder
Attention
Dot products
Decoder
<p> <p> <s> Sie stimmen zu
Sie stimmen zu </s>

Model Overview

01

Encoder

Simple overview of the encoder.

02

Decoder

Simple overview of the decoder.

03

Attention

Simple overview of the attention mechanism.

04

Results

Walk through the results.

01

ABOUT

The Encoder

Digests the source sentence for the decoder.

The Encoder

01

Embedding

Word Embedding + Positional Embedding.

02

Convolution

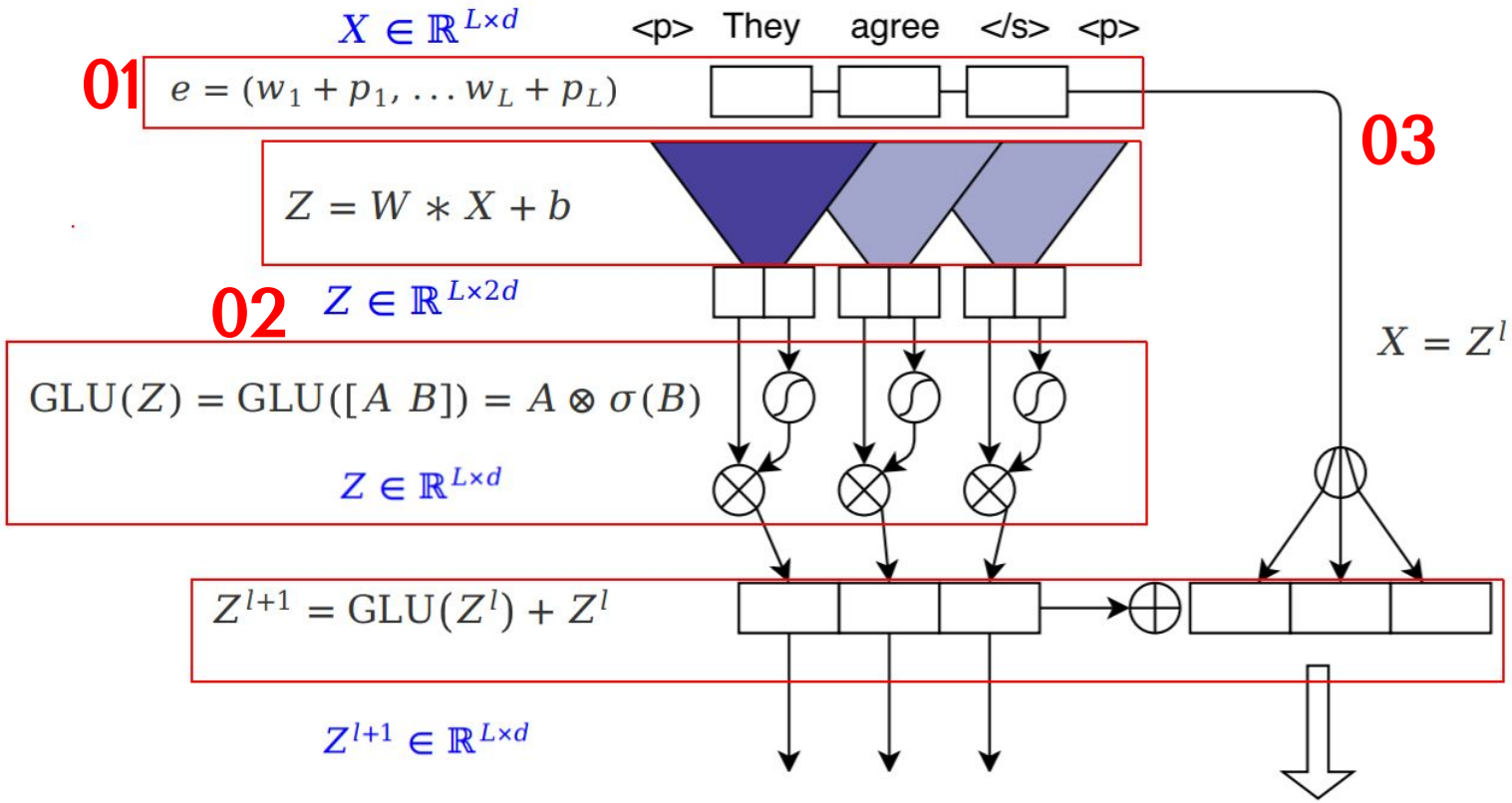
1D Convolution + GLU activation function.

03

Residual

Residual connections are applied.

Encoder


$X \in \mathbb{R}^{L \times d}$
<p> They agree </s> <p>
01
$e = (w_1 + p_1, \ldots w_L + p_L)$
$Z = W * X + b$
02
$Z \in \mathbb{R}^{L \times 2d}$
$\mathrm{GLU}(Z) = \mathrm{GLU}([A\ B]) = A \otimes \sigma(B)$
$Z \in \mathbb{R}^{L \times d}$
03
$X = Z^l$
$Z^{l+1} = \mathrm{GLU}(Z^l) + Z^l$
$Z^{l+1} \in \mathbb{R}^{L \times d}$

02

ABOUT

The Decoder

Digests the target sentence along with the encoder’s digestion.

The Decoder

01

Embedding

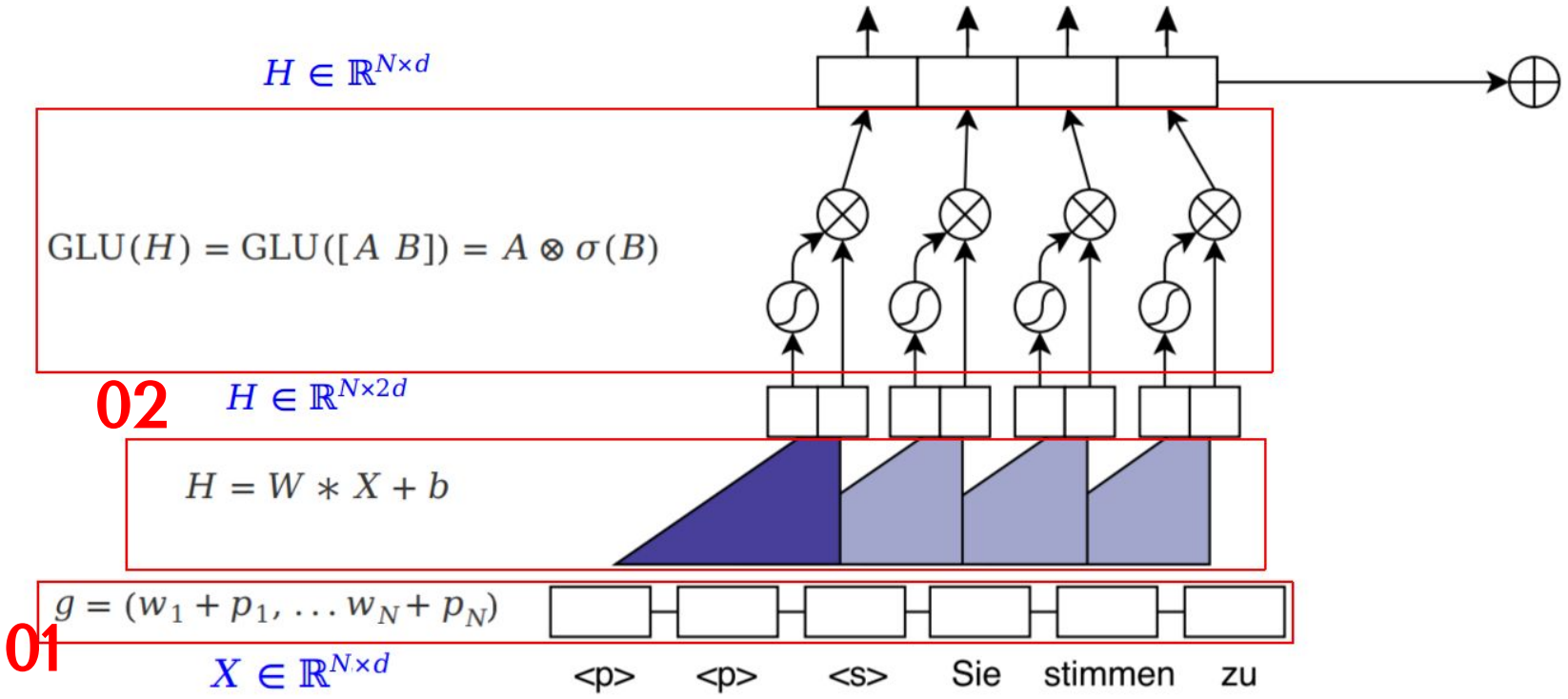
Word Embedding +
Positional Embedding.

02

Convolution

1D Convolution + GLU
activation function.

Decoder


$H \in \mathbb{R}^{N \times d}$
$\mathrm{GLU}(H) = \mathrm{GLU}([A\ B]) = A \otimes \sigma(B)$
02
$H \in \mathbb{R}^{N \times 2d}$
$H = W * X + b$
$g = (w_1 + p_1, \ldots w_N + p_N)$
01
$X \in \mathbb{R}^{N \times d}$
<p>
<p>
<s>
Sie
stimmen
zu

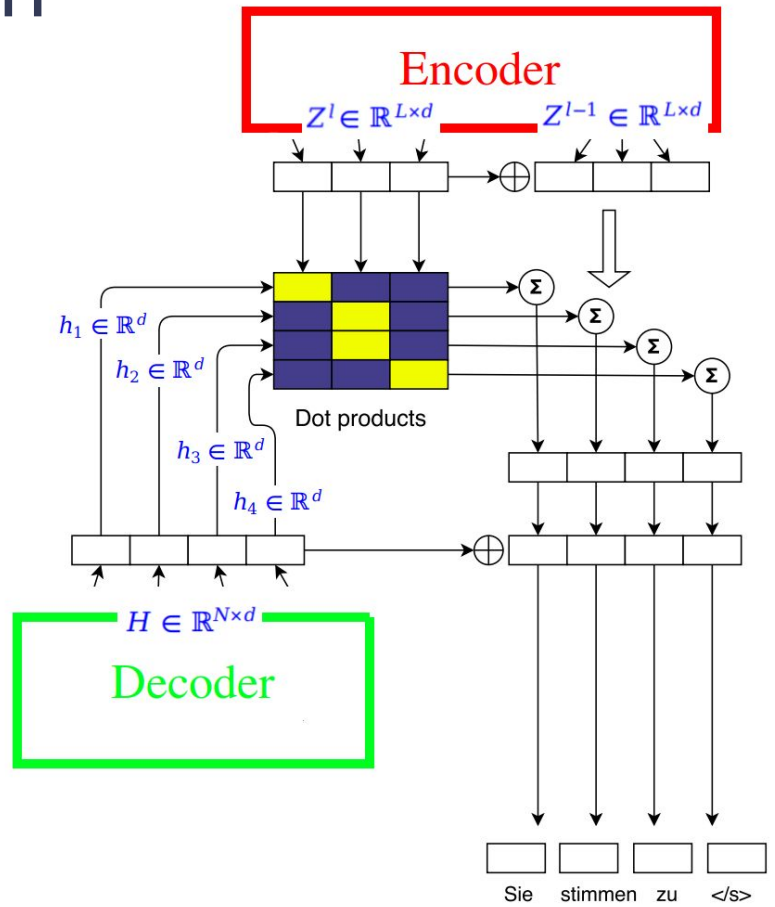
03

ABOUT

Attention Mechanism

Decides which part of the decoder is related to the encoder by how much.

Attention

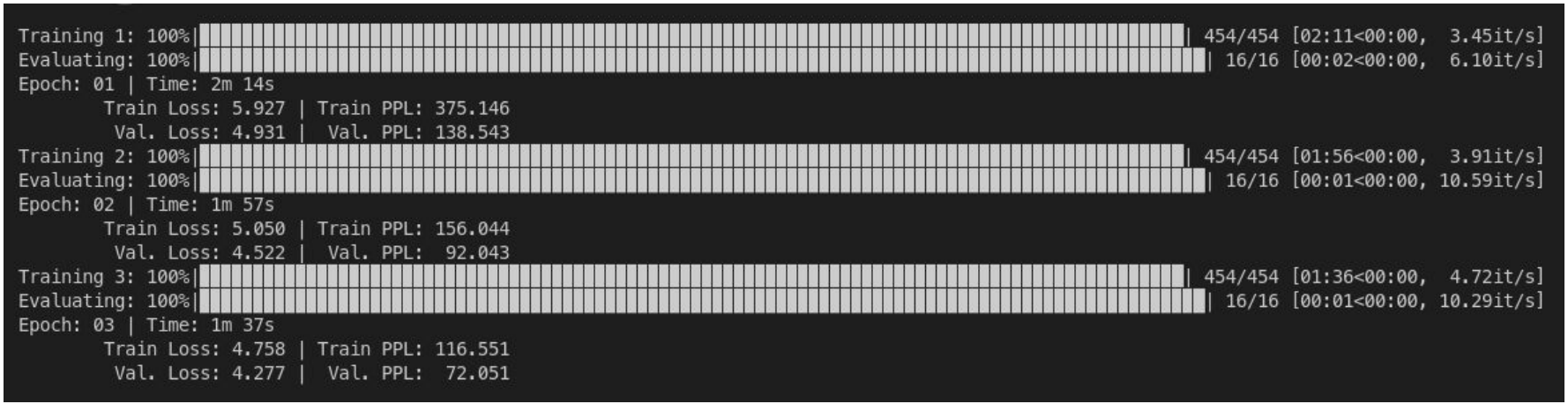

Encoder
$Z^l \in \mathbb{R}^{L \times d}$
$Z^{l-1} \in \mathbb{R}^{L \times d}$
$h_1 \in \mathbb{R}^d$
$h_2 \in \mathbb{R}^d$
$h_3 \in \mathbb{R}^d$
$h_4 \in \mathbb{R}^d$
Dot products
$H \in \mathbb{R}^{N \times d}$
Decoder
Sie
stimmen
zu
</s>

04

ABOUT Results

Showcase my implementation

Showcasing


Training 1: 100%| 454/454 [02:11<00:00, 3.45it/s]
Evaluating: 100%| 16/16 [00:02<00:00, 6.10it/s]
Epoch: 01 | Time: 2m 14s
Train Loss: 5.927 | Train PPL: 375.146
Val. Loss: 4.931 | Val. PPL: 138.543
Training 2: 100%| 454/454 [01:56<00:00, 3.91it/s]
Evaluating: 100%| 16/16 [00:01<00:00, 10.59it/s]
Epoch: 02 | Time: 1m 57s
Train Loss: 5.050 | Train PPL: 156.044
Val. Loss: 4.522 | Val. PPL: 92.043
Training 3: 100%| 454/454 [01:36<00:00, 4.72it/s]
Evaluating: 100%| 16/16 [00:01<00:00, 10.29it/s]
Epoch: 03 | Time: 1m 37s
Train Loss: 4.758 | Train PPL: 116.551
Val. Loss: 4.277 | Val. PPL: 72.051

THANKS

Do you have any questions?

Code: https://github.com/Anwarvic/ConvS2S