<p> <p> X_1 X_2 X_3 <EOS>   <EOS> Y_1 Y_2 <p> <p>

Embedding

kernel_width=3

Convolutional

GLU Layer

⊕

⊕

Attention Matrix

Y_1' Y_2' <EOS>

<p> <p> X_1 X_2 X_3 <EOS>

<EOS> Y_1 Y_2 <p> <p>

Embedding

kernel_width=3

Convolutional

GLU Layer

$\oplus$

Attention
Matrix

$\oplus$

Y_1' Y_2' <EOS>

Length Embedding

Embedding

<EOS> Y_1 Y_2 <p>   <p>

⊗   ⊗   ⊗   ⊗   ⊗

kernel_width=3

Convolutional

GLU Layer

Attention
Matrix

⊕

Y'_1 Y'_2 <EOS>

Embedding Convolutional GLU Layer Attention Matrix