

Transformer Encoder

$x_{0,0}^O$
 $x_{1,0}^O$
 \vdots
 $x_{i,j}^O$
 \vdots

OA
Encoder

h^O

H^O

SelfAttn

softmax

A^O

λ^O

$x_{0,0}^I$
 $x_{1,0}^I$
 \vdots
 $x_{i,j}^I$
 \vdots

IS
Encoder

H^I

SelfAttn

A^I

λ^I

h^I

Transformer Decoder

Self-Attention

ADD & Normalize

Enc-
Dec Att

$\tilde{\lambda}^O$

Σ

$\tilde{\lambda}^I$

Enc-
Dec Att

ADD & Normalize

Feed Forward

ADD & Normalize

Summary

y