

ASR Block

Input: X_1, X_2, \dots, X_t

SSL
Encoder

Conformer
Encoder

Concat

Detach

Attention
Encoder

Linear
Projection

Attention
Decoder

CTC
Decoder

Text: W_1, W_2, \dots, W_k

DID Block

CTC
Softmax

Bottleneck
Encoder

RoBERTa
Encoder

Gating

Attention
Encoder

Linear
Projection

Mean
Pooling

Classifier

DID: D_1, D_2, \dots, D_m

