

## **Model Architecture**

LAS standard architecture

LSTM layer: 1

LSTM hidden dimension: 256

Pyramid LSTM layer: 3

Pyramid LSTM hidden dimension: 256

Attention value size: 256

Attention key size: 256

Decoder layer: 2

Decoder hidden dimension: 256

## **Loss Function**

Cross entropy loss + Adam

## **Hyper Parameters**

Batch size: 64

Learning rate:  $5e-3$

Weight decay:  $5e-5$

Scheduler: cosine annealing schedule with maximum number of iteration 120

Number of epoch: 200, highest score achieving at 176