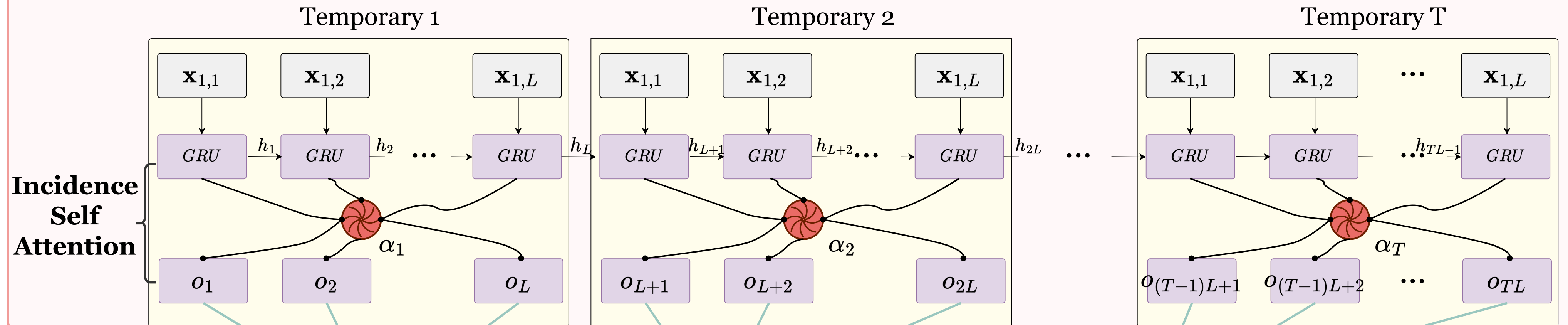


# Encoder

## Incidence self attention Encoder

$\alpha_i$  is Incidence Self Attention weight tensor



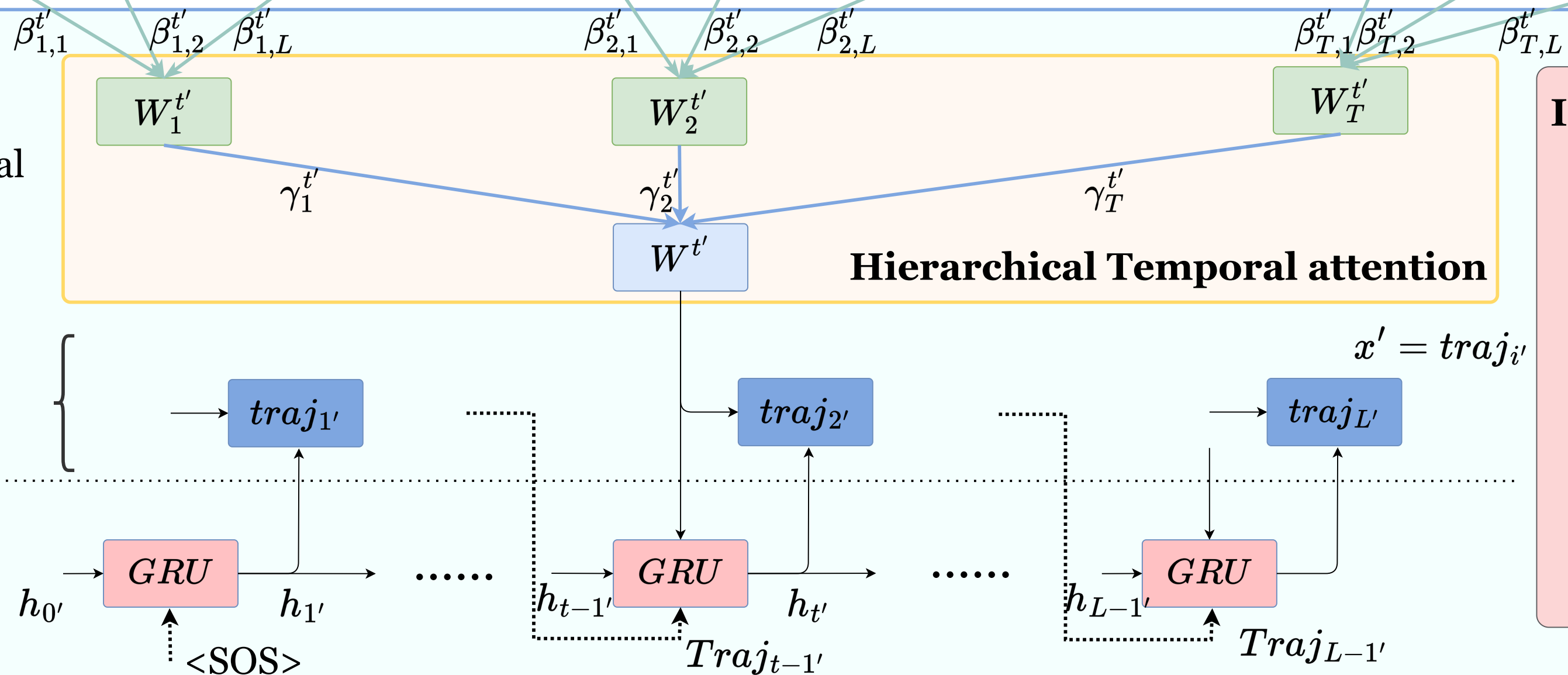
# Decoder

# A hierarchical temporal attention-based GRU decoder

## Output

Concat Linear and  
Softmax

**GRU** {  
Gated  
Recurrent Unit



## In Train Phase Only

## Cross-Entropy Loss

