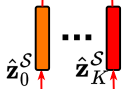
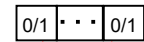
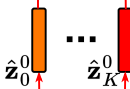
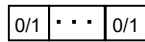


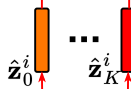
Scene Predictions



Char-1 Predictions



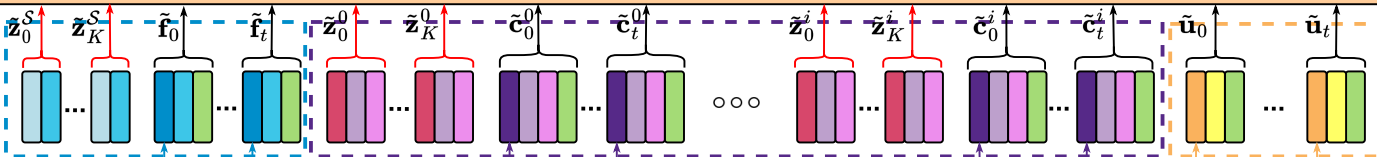
Char-P Predictions



Shared

Shared

Transformer Encoder Layer (x2)



Linear Layer

Linear Layer

Linear Layer

MViT (Kinetics 400)

ResNet50 (FER 2013)

Ft. RoBERTa

