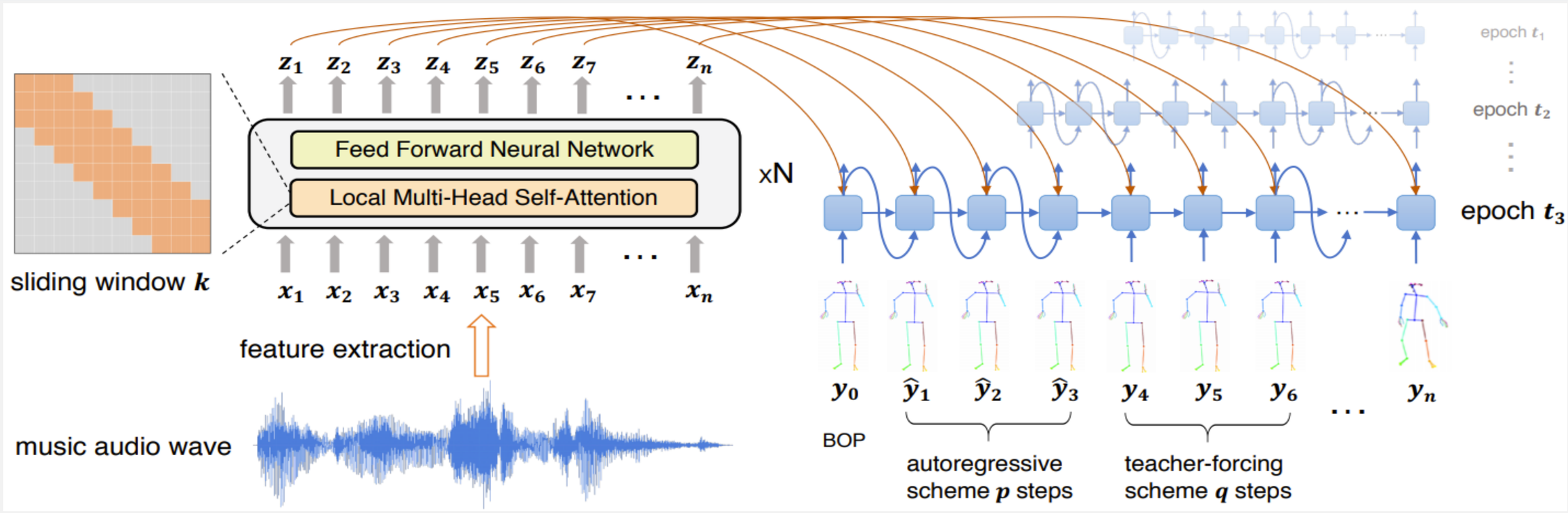
Model Architecture



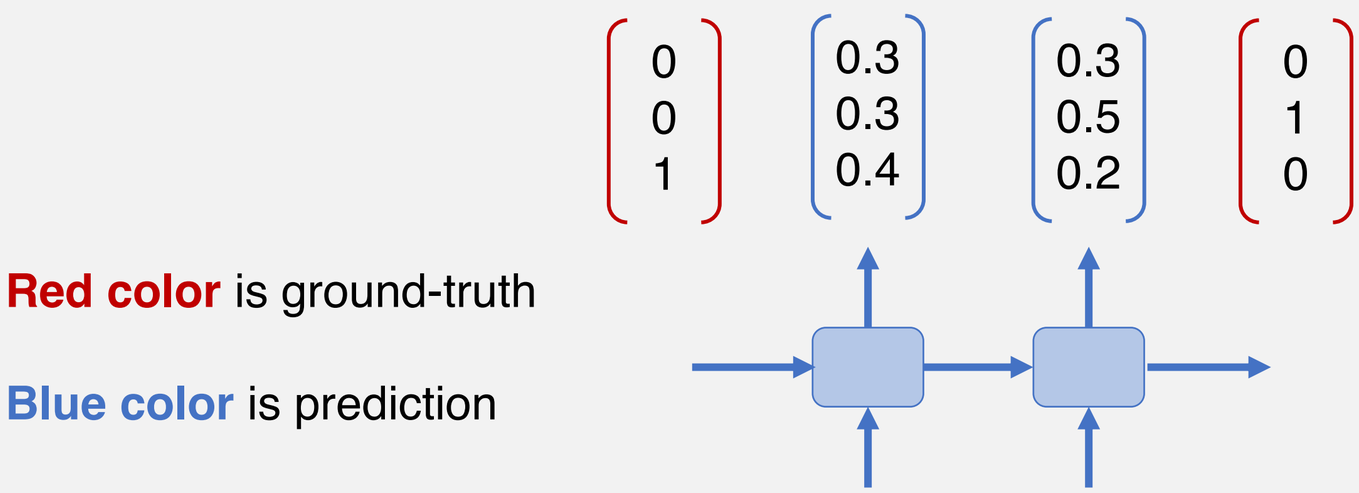
$$p = f(t) \in \{const, \lfloor \lambda t \rfloor, \lfloor \lambda t^2 \rfloor, \lfloor \lambda e^t \rfloor\}$$

Dynamic Auto-Condition Learning Approach

Dance is a long-term sequence composed of human motions with the spatio-temporal structure. Hence, generating dances also suffers from the exposure bias issue known in NLG, referring to the train-test discrepancy of autoregressive models. This issue would quickly accumulate the prediction errors at inference and thus make the generated motion sequences rapidly converge to the mean poses, i.e., “freezing motions”.

Scheduled Sampling v.s. Dynamic Auto-Condition Learning Approach

- The former is proposed for the sequence generation of symbols in discrete space while the latter is tailored for the long sequence generation of real-valued vectors in continuous space (each motion is represented by a dozen of 2D or 3D keyjoints).
- The bias of predicted probability distribution over vocabulary can be corrected by sampling strategy at inference. While any small biases of predicted motions (i.e., motion vectors) at each time-step will be accumulated and propagated to the future.



$$prediction\ bias = (|0\ - 0.3| + |0 - 0.3| + |1 - 0.4|) + (|0 - 0.3| + |1 - 0.5| + |0 - 0.2|)$$