

Supplementary Material

MambaMIL: Enhancing Long Sequence Modeling with Sequence Reordering in Computational Pathology

The Hong Kong University of Science and Technology

Algorithm 1 SR-Mamba Block Process

Input: instance sequence $X_{l-1} : (B, M, D)$
Output: instance sequence $X_l : (B, M, D)$
B: batch size, M: instance number, D: dimension
 $X'_{l-1} : (B, M, D) \leftarrow \text{LayerNorm}(X_{l-1}), \quad z : (B, M, E) \leftarrow \text{Linear}^z(X'_{l-1})$
Original Sequence: os
 $x_{os} : (B, M, E) \leftarrow \text{Linear}^{x0}(X'_{l-1})$
Reordered Sequence: rs
 $x_{rs} : (B, M, E) \leftarrow \text{Reordering}(\text{Linear}^{x1}(X'_{l-1}))$
process sequences with distinct orderings
for o in {os, rs} **do**
 $x_o : (B, M, E) \leftarrow \text{SiLU}(\text{Conv1d}_o(x))$
 $B_o : (B, M, N) \leftarrow \text{Linear}^B(x_o), \quad C_o : (B, M, N) \leftarrow \text{Linear}^C(x_o)$
 $\Delta_o : (B, M, E) \leftarrow \log(1 + \exp(\text{Linear}^A(x_o) + \text{Parameter}^A))$
 $A_o : (B, M, E, N) \leftarrow \Delta_o \otimes \text{Parameter}^A, \quad B_o : (B, M, E, N) \leftarrow \Delta_o \otimes B_o$
 $y_o : (B, M, E) \leftarrow \text{SSM}(A_o, B_o, C_o)(x_o)$
end for
 $y_{os} : (B, M, E) \leftarrow y_{os} \odot \text{SiLU}(z), \quad y_{rs} : (B, M, E) \leftarrow y_{rs} \odot \text{SiLU}(z)$
residual connection
 $X_l : (B, M, D) \leftarrow \text{Linear}(y_{os} + y_{rs}) + X_{l-1}$
return X_l

Table 1: Hyper-parameter configurations.

<i>Datasets</i>	BLCA	BRCA	COADREAD	KIRC	KIRP	LUAD	STAD	BRACS	NSCLC
Sample	437	1023	572	498	261	455	363	545	1053
Min Length	414	283	30	319	383	85	26	49	85
Max Length	34174	36618	27418	30679	62235	45785	26130	26600	45785
Average Length	14419	8893	6953	12121	12424	9897	9402	7812	10515
Learning Rate	2e-4	2e-5	2e-5	2e-4	2e-4	2e-4	2e-4	1e-5	2e-5
Segments Size R	5	5	10	10	10	5	5	10	5