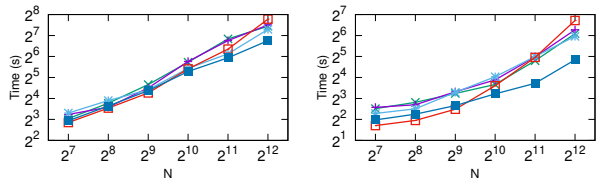
Table 4: Hyperparameters

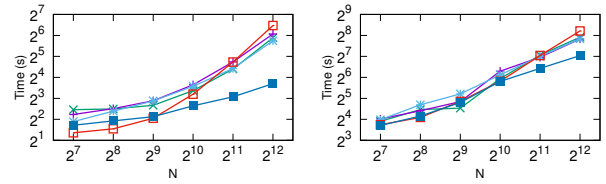
Dataset name	Batch size	Epochs
BasicMotions	32	100
Cricket	4	200
UWaveGestureLibrary	32	100
EigenWorms	4	100
Epilepsy	4	200
Handwriting	4	200
RacketSports	32	200
Libras	4	100
SelfRegulationSCP2	4	200
SelfRegulationSCP1	64	100
ArticularyWordRecognition	64	200
CharacterTrajectories	8	200
Phoneme	64	100
SpokenArabicDigits	4	200
Synthetic A	8	200
Synthetic B	8	200



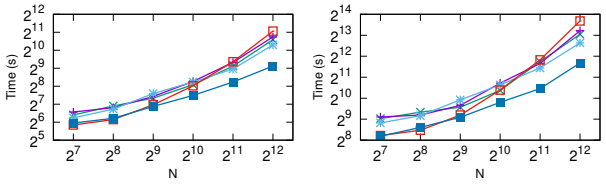
(a) EigenWorms (b) Epilepsy

(c) Handwriting (d) RacketSports

Figure 16: Other results of training time (Part A)



(a) Libras (b) SelfRegulationSCP2



(c) Phoneme (d) Synthetic B

Figure 17: Other results of training time (Part B)

Table 3: Details of datasets

Dataset name	No.Instance	Length
BasicMotions	40	100
Cricket	108	1197
UWaveGestureLibrary	120	315
EigenWorms	128	17984
Epilepsy	137	206
Handwriting	150	152
RacketSports	151	30
Libras	180	45
SelfRegulationSCP2	200	1152
SelfRegulationSCP1	268	896
ArticularyWordRecognition	275	144
CharacterTrajectories	1422	1422
Phoneme	6599	3315
SpokenArabicDigits	6599	6599
Synthetic A	10000	1024
Synthetic B	20000	1024

A APPENDIX

A.1 Details of datasets

We provide the size of training set (instance number) and time series length for the 14 real datasets and 2 synthetic datasets shown in Table 3. Readers may refer to the UEA archive paper [1] for more information about 14 datasets. For Synthetic A and B, we use numpy to randomly generate them.

A.2 All results of training time

We provide the experimental results on the efficiency of DARKER for the rest 8 datasets not presented in Section 5.3. As shown in Figure 16 and Figure 17, we observed that the trends for these datasets are consistent with the experimental results in Section 5.3, and DARKER continues to be 3×-4× faster than vanilla transformer and 1.5×-3× faster than other SOTAs for long sequences.

A.3 Details of hyperparameters

For training the transformer model, we follow the previous work [54, 58] to split the training set into two parts, namely 80% and 20%, where the 20% part is as the validation set to tune the hyperparameters. Then, the model is trained on the whole training set based on the tuned hyperparameters. The details of hyperparameters of all datasets are shown in Table 4.