**Reply to Reviewer #1**

Though there is no need to adjust our manuscript after the first revision, we would like to answer the questions of one reviewer here as anyone interested in this work may also have these questions.

*Problem #1*

*In Fig.6, the complexity of FPPformer with or without decoder is analogous, which is irrational.*

Response #1

The curves of FPPformer-Enc in Fig. 6(b) are not correct, albeit that they do not influence the final conclusion. The true curves are shown below.
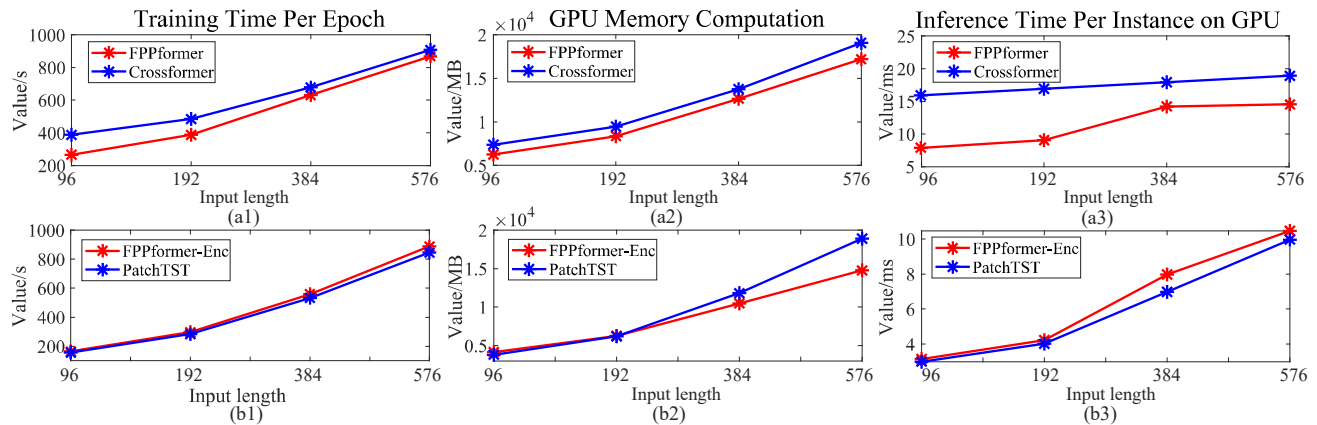


Fig. 6. The results of training time per epoch (1), GPU memory computation (2) and inference time per instance (3) in two experiments (a)(b).

*Problem #2*

*Please specify the legends in Fig. 11.*

Response #2

The four blocks with different colors and numbers refer to the patch positions where the highest attention scores locate at. A bigger number indicates a temporally later patch position.
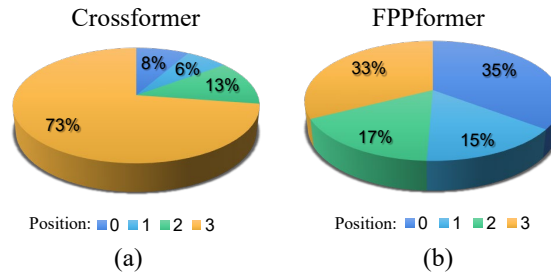


Fig. 11. The distribution of the highest cross-attention score positions of the third stage in Crossformer (a) and the first stage in FPPformer (b).

-----End of Reply to Reviewer #1------