R1

Q1. *On page 4, line 404, you mention the issue of overfitting with the local reconstruction. Why this unreliable local reconstruction could be applied to calculate the anomaly score (Eq. 8)?*

Reply: The local reconstruction $\hat{X}_L$ mainly depends on local context, which can smooth the noise (white noise). However, the anomalous data have abrupt changes in large segments or large amplitude, and $\hat{X}_L$ can not smooth them out. As a result, the $\hat{X}_L$ will overfit anomalous data. In our model, $\hat{X}_G$ can smooth the anomalous data benefiting from its dependence on remote context. If we use $|\hat{X}_G - X|$ to detect anomalies, the noise in X will impact the accuracy, especially for noise data. However, $|\hat{X}_G - \hat{X}_L|$ can reduce the impact of noise, owing to the local smoothing power of $\hat{X}_L$. This working mechanism can be illustrated in Fig. 1 and Fig. 8.

W1.*The novelty is relatively limited -- detecting anomalies from both local and global perspectives has been extensively studied in existing works [1].*

References: [1] Unsupervised Anomaly Detection in Time-series: An Extensive Evaluation and Analysis of State-of-the-art Methods, Arxiv, 2022

Reply: In the provided reference[1], their local and global perspectives refer to the type of point-by-point anomalies: global outliers and contextual outliers (Lai 2021). However, our local and global refer to the method of reconstruction. Ideally, the local reconstruction $\hat{X}_L$ can smooth noise by its local context, and the global reconstruction $\hat{X}_G$ can smooth both noise and **anomaly** by its global context. $|\hat{X}_G - \hat{X}_L|$ can become large in the anomaly area but not affected by noise in the noise area, because both $\hat{X}_G$ and $\hat{X}_L$ can smooth out noise. To the best of our knowledge, this is the first method to detect anomalies by the deviation of local and global reconstructions.
Kwei-Herng Lai, Daochen Zha, Junjie Xu, Yue Zhao, Guanchu Wang, and Xia Hu. 2021. Revisiting Time Series Outlier Detection: Definitions and Benchmarks. In Proceedings of the Neural Information Processing Systems Track on Datasets and Benchmarks, J. Vanschoren and S. Yeung (Eds.), Vol. 1. Curran.

W2. *The authors state that local reconstruction enhances robustness to noise, yet there appears to be no experimental evidence provided to back this claim. Moreover, in Figure 4 on page 7, the local reconstruction error is large. This seems to conflict with the earlier assertion.*

Reply:

- The local reconstruction enhances robustness to noise attributed to the following reasons. (1) $\hat{X}_G$ smooth anomaly data by remote context and recover normal data. If

we use $|\hat{X}_G - X|$ as the anomaly score in the noise data, the false alarm may be generated by the noise in X. (2) However, if we use the $\hat{X}_L$ in the place of X and use $|\hat{X}_G - \hat{X}_L|$ as the anomaly score, the impact of noise will be reduced. Because both $\hat{X}_G$ and $\hat{X}_L$ can smooth out noise.

- In Figure 4, we analyze the model's working mechanism for three types of anomaly scores, $|\hat{X}_G - \hat{X}_L|$, $|\hat{X}_G - X|$, and $|\hat{X}_L - X|$. we use $|\hat{X}_G - \hat{X}_L|$ as the anomaly score because it can alleviate the impact of noise by local- and global- smoothing, and keep structural change by $\hat{X}_L$. The experimental results demonstrate that it can achieve the best performance on most datasets. The **local reconstruction error** $|\hat{X}_L - X|$ is large because the local reconstruction $\hat{X}_L$ can only smooth noise by local context but not for large abrupt changes. As a result, its F1 score is very low in almost all datasets. These results support our assumptions.

- For more noise experiments, we added noise to the test set with signal-to-noise ratios ranging from 10dB to 100dB. A higher signal-to-noise ratio indicates less noise, and "None" indicates no noise added. To verify the contribution of local reconstruction $\hat{X}_L$ to the noise robustness performance , we experimentally compare the F1 scores using $MSE\left(\hat{X}_L, \hat{X}_G\right)$ with $MSE\left(X, \hat{X}_G\right)$ and report the results in the Figure 1. In the univariate dataset UCR and the multivariate dataset SWaT and PSM, $MSE\left(X, \hat{X}_G\right)$ is weaker than $MSE\left(\hat{X}_L, \hat{X}_G\right)$. In the SWaT dataset, when the noise becomes 10 dB, the detection performance of $MSE\left(X, \hat{X}_G\right)$ decreases sharply, while $MSE\left(\hat{X}_L, \hat{X}_G\right)$ still maintains a more stable level. The above results indicate that using $\hat{X}_L$ in the place of X can boost the robustness to noise.
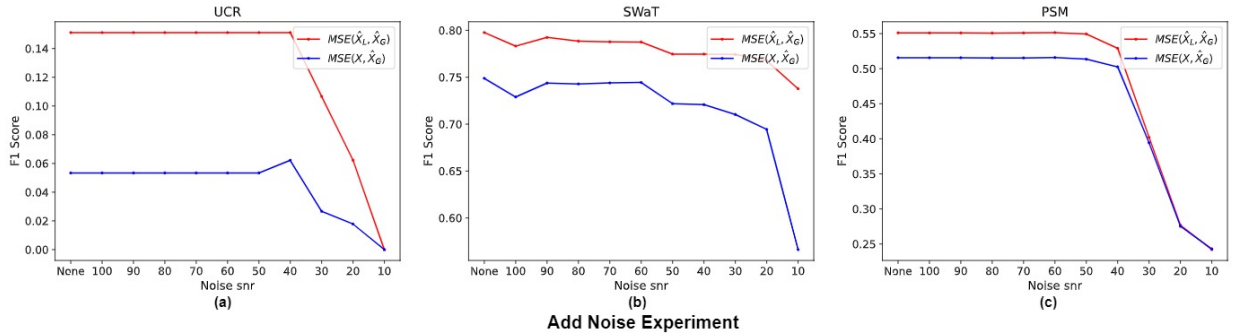

Figure 1. Robustness of Local Reconstruction on noise

W3. *The experimental results presented by the authors are inconsistent with those reported in previous research, which weakens the Technical Quality of this work.*

Reply: The inconsistencies between the experimental results provided by MGRD and those by previous studies are summarised in the following reasons.
(1) **Dataset**: Figure 2 shows the data information provided in the different baselines. The anomaly ratios in MGRD and Anomaly Transformer(AT) are highly overlapping, and some of the dataset information is slightly different from that published in methods such as GDN, TranAD, etc. (The inconsistency of the data information is common, e.g., the anomaly ratio

of the MSL dataset is different in the OmniAnomaly, MTAD-GAT, and AT are published differently).

	OmniAnomaly	GDN	MTAD-GAT	TranAD	AT	MGRD
MSL	10.72%	-	10.27%	10.72%	10.5%	10.53%
PSM	-	-	-	-	27.8%	27.76%
SMAP	13.13%	-	13.13%	13.13%	12.8%	12.79%
SMD	4.16%	-	-	4.16%	4.2%	4.21%
SWaT	-	11.97%	-	11.98%	-	12.14%
WADI	-	5.99%	-	5.99%	-	5.77%

Information statistics on data

Figure 2. the static of anomaly rates in different papers

(2) **Data Preprocessing.** On some datasets, Anomaly Transformer used the test set as validation data in the published code. However, we divided that dataset into the train/validation/test sets during the experimental process and applied them to all the compared methods. MTAD-GAT carried out the two processes of data normalization and data cleaning during the data processing process, while the MGRD did not use the data cleaning process for all models to validate their robustness to noise.

(3) **Point-Adjust Strategy**. In this paper, we use the experimental results without the point-adjust strategy as a guideline to avoid overestimation of detection performance[11].

(4) **Threshold**. The threshold is very sensitive to many baselines. This work chooses the real anomaly ratio as the threshold for all models' anomaly detection. The unified threshold-choosing method can reduce the sensitivity of the threshold and concentrate on anomaly score modeling.

(5) **Code implementation**. GDN and MTAD-GAT are implemented using GraphAn. MTAD-GAT uses the implementation in TranAD's published code. In the data window division process, the window size is equal to the number of data features.

W4. *The code necessary for verifying the experiments hasn't been provided.*

Reply: We provide an anonymous URL for code. https://anonymous.4open.science/r/MGRD