PROJECT PAGE and APPENDIX: https://zhaobenyun.github.io/CUBIT/

Due to space limitation, we prepare a detailed reply to reviewers at:

28E4"Importance":
Our paper focuses on small model of single-stage object detection algorithms suitable for real-time defect detection on upper-computers, where inference speed and model size are crucial. While transformer-based models are indeed powerful, their larger size doesn't meet our criteria for real-time processing.
About loss function, we don't make any other adjustments, just using the standard loss function in official code of each algorithm.

28E4"Experimental" & 04B2"Additional comments(2)":
In Section3.2, we explained the selection of YOLOv6-nano as baseline for its great balance between inference speed, model size, and accuracy. After adding our 'GIPFPP' module, YOLOv6-nano greatly improved in detection accuracy and model size, which aligns with our goal of lightweight yet effective model for further real-world applications.
mAP50:95 is a more stringent metric than mAP50. Original YOLOv6-nano is much smaller and faster than YOLOv6-small, and YOLOv6-nano+GIPFPP shows greater detection ability than YOLOv6-small in mAP50:95, so it is no need to compare more.
We also applied 'GIPFPP' to YOLOv5-nano. After comparing YOLOv5-nano with YOLOv6-nano, we found that YOLOv6-nano was better. We only present the results of YOLOv6-nano in this paper and put more experiment results in APPENDIX.

28E4"Presentation" & 04B2"Additional comments(1)" &1745"Additional comments(4)":
For texts in the figures are small, we have adjusted. In the new version paper, the figures and texts will be clearly visible. In PROJECT PAGE, we also present important figures in a higher resolution.

04B2"Novelty" & 04B2"Additional comments(1)" & 1745"Additional comments(3)":
The focus of our paper is to present a specialized dataset 'CUBIT', which is the FIRST high-resolution defect detection dataset, covering multiple infrastructures. Fig2 is necessary to present the numerical analysis of CUBIT.
In previous ICASSP, there are several datasets in different engineering fields. Since datasets are very critical work in deep learning. We did more than simply present a professional dataset, we also provided many benchmark results.
Due to the large amount of data, 'CUBIT' is preparing to be open-source and some samples are uploaded. The full dataset will come after acceptance.

1745"Additional comments(1)":
The definition of target size was explained in notes under Fig2.

1745"Additional comments(2)":
About preprocessing, when the images are input into model, model directly resizes images to 1024, without other skills, because this part isn't the technical focus of our model improvement. If using too many preprocessing skills to improve detection ability, it will be an unfair comparison.

Reviewer 28E4:

Importance/Relevance to ICASSP 2024:

Our paper focuses on small model of single-stage object detection algorithms suitable for real-time defect detection on upper-computers, where inference speed and model size are crucial. Since our final goal is to deploy our model onto the onboard computer of unmanned robots, such as UAV and UGV.

While transformer-based models are indeed powerful, their larger size doesn't meet our criteria. Since we are aiming at achieving real-world real-time infrastructures defects detection.