Dear Technical Program Committee and Reviewers,

Thank you for your insightful comments. We would like to address your concern regarding our paper. I will answer one by one according to the serial number and questions of the reviewers.

Reviewer 28E4:

'Importance/Relevance to ICASSP 2024':

About model comparison, 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.

About objective loss function, we do not make any other adjustments about this part. We just used the standard loss function in official repository from GitHub of each algorithm. And the loss function of object detection task contains three parts: object loss, classification loss and confidence loss.

'Experimental Validation':

In Section3.2, we explained the selection of YOLOv6-nano as baseline for its great balance between inference speed, model size, and accuracy. YOLOv6-nano is much smaller and faster than YOLOv6-small, so we do not choose YOLOv6-small as baseline. 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. YOLOv6-nano+GIPFPP shows greater detection ability than YOLOv6-small under mAP50:95, and YOLOv6-nano-GIPFPP is smaller than original YOLOv6-nano. Therefore, whether it is model size, inference speed, or detection accuracy, YOLOv6-nano-GIPFPP is better than YOLOv6-small, which has met our needs, and we do not need to improve the YOLOv6-small model which is many times larger than YOLOv6-nano-GIPFPP.

Except YOLOv6, 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.

'Clarity of Presentation':

We have modified the texts in each Figure. 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.

Reviewer 04B2:

'Novelty/Originality':

By reading the previous ICASSP papers, we found that there are also many specialized datasets in different engineering fields. Since datasets are very critical work in deep learning related works. More importantly, we did more than simply present a professional dataset, we also provided many benchmark results to researchers for further application and research.

Additional Comments to author (1):
The focus of our paper is to present a specialized dataset 'CUBIT', which is the FIRST high-resolution defect detection dataset, covering multiple infrastructures, especially buildings. Fig2 is necessary to present the numerical analysis of CUBIT dataset, this part very common and essential in papers about dataset.

Additional Comments to author (2):
The focus of model improvement is to make the model more lightweight without sacrificing detection ability. We think that simple methods that show good results are meaningful. Our proposed module 'GIPFPP' simply upgrades from common SPPF and CSPSPPF modules. And when adding GIPFPP to baseline model YOLOv6-n, our desired results have achieved: a 10% reduction in model size and a 3% improvement in accuracy on the strict metric mAP50:95.

Additional Comments to author (3):
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.

Reviewer 1745:

Additional Comments to author (1):
The definition of target size was explained in notes under Fig2. If the paper is accepted and can be revised later, I will add this sentence to the corresponding paragraph.

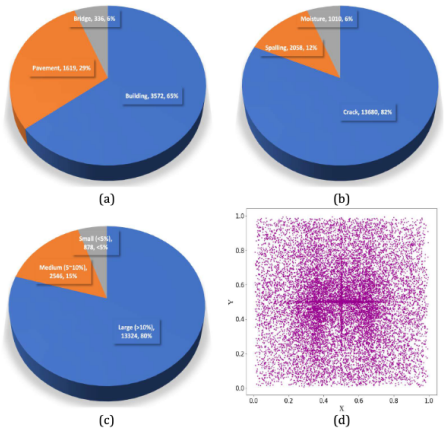

Fig. 2. (a) Defect Collection Scenarios; (b) Defect Categories; (c) Defect Target Dimensions: Large targets are exceeding 10% of image dimensions, medium-size targets are ranging from 5% to 10% and small targets are less than 5%. **(d) Defect Target Distribution:** Each scatter represents the position of one target relative to the center of the image.

Additional Comments to author (2):
Fig 2(d) shows the location distribution of all defect targets in CUBIT dataset. However, not all kinds of defects follow this distribution, especially moisture, because moisture are relatively few, so it can not follow this distribution completely.

Additional Comments to author (3):
About image preprocessing part, when the images are input into model, model directly resizes images to 1024x1024, without other skills, because this part isn't the technical focus

of our model improvement. If we use many preprocessing skills to improve detection ability, the strengthen of our GIPFPP module will be reduced, and overly complex data preprocessing will slow down the reasoning speed of the model.

Additional Comments to author (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.

Best regards,
Benyun Zhao
Unmanned Systems Research Group
The Chinese University of Hong Kong