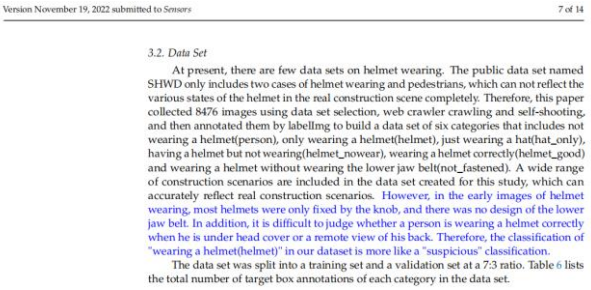
Response to Reviewer 1 Comments

Point 1: As shown in Figure 3, there are some logical relationships between the proposed six cases. For instance, wearing a helmet correctly and wearing a helmet without wearing the lower jaw belt both belong to the case of wearing a helmet. Should the six cases be reduced to five cases by removing that of "wearing a helmet"?

Response 1: Thank you for the comment. The answers are as follows: most of the existing helmet-wearing detection methods only carry out 2-classification experiments on data sets SHWD, which detect whether to wear a helmet or not. We hope to refine the study by more subdividing the states of the helmet wearing while improving the algorithm. In the initial sub classification of the data set, we only considered five situations: not wearing a helmet(person), wearing a helmet, just wearing a hat(hat_only), having a helmet but not wearing(helmet_nowear), wearing a helmet correctly(helmet_good), and wearing a helmet without wearing the lower jaw belt(not_fastened). However, in the actual research process, we found that in the early images of helmet wearing, most helmets were only fixed by the knob, and there was no design of the lower jaw belt when we collected images through web crawlers, public dataset selection and other methods. In addition, it is difficult to judge whether a person is wearing a helmet correctly when he is under head cover or a remote view of his back. Therefore, the classification of "wearing a helmet(helmet)" in our dataset is more like a "suspicious" classification. For this problem, we explained it in Chapter 3.2 from Line 174 to Line 178 in Page 7 of the manuscript to avoid the same misunderstanding. Thanks again for your reminder. The change has been highlighted in the revised manuscript.



Version November 19, 2022 submitted to *Sensors* 7 of 14

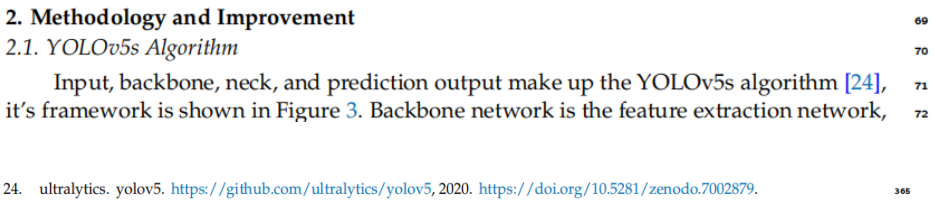
3.2. Data Set

At present, there are few data sets on helmet wearing. The public data set named SHWD only includes two cases of helmet wearing and pedestrians, which can not reflect the various states of the helmet in the real construction scene completely. Therefore, this paper collected 8476 images using data set selection, web crawler crawling and self-shooting, and then annotated them by labelImg to build a data set of six categories that includes not wearing a helmet(person), only wearing a helmet(helmet), just wearing a hat(hat_only), having a helmet but not wearing(helmet_nowear), wearing a helmet correctly(helmet_good) and wearing a helmet without wearing the lower jaw belt(not_fastened). A wide range of construction scenarios are included in the data set created for this study, which can accurately reflect real construction scenarios. However, in the early images of helmet wearing, most helmets were only fixed by the knob, and there was no design of the lower jaw belt. In addition, it is difficult to judge whether a person is wearing a helmet correctly when he is under head cover or a remote view of his back. Therefore, the classification of "wearing a helmet(helmet)" in our dataset is more like a "suspicious" classification.

The data set was split into a training set and a validation set at a 7:3 ratio. Table 6 lists the total number of target box annotations of each category in the data set.

Point 2: The YOLOv5s algorithm should be properly cited, especially in Section 2.1 (there is no citation in the manuscript).

Response 2: Thank you for the comment. The answers are as follows: In the process of writing this paper, we also noticed this problem. The YOLOv5 algorithm is released by a particle physics and artificial intelligence startup company. Up to now, the authors of YOLOv5 have not published papers, so there is no way to quote the papers. But we cited the github address of YOLOv5 algorithm(https://github.com/ultralytics/yolov5). The added parts are the 71th line in page 3 and the 24th reference in line 365 in page 14.



2. Methodology and Improvement

2.1. YOLOv5s Algorithm

Input, backbone, neck, and prediction output make up the YOLOv5s algorithm [24], it's framework is shown in Figure 3. Backbone network is the feature extraction network,

24. ultralytics. yolov5. https://github.com/ultralytics/yolov5, 2020. https://doi.org/10.5281/zenodo.7002879.