Data Processing Flow Description

The data processing flow includes dataset splitting, data augmentation, and random seed management, aiming to provide consistent and high-quality input data for model training. The following are the specific steps of the data processing flow:

Dataset Splitting

Our dataset is first split according to a predetermined ratio into training, validation, and test sets. During the splitting process, fixed random seeds are used to ensure the reproducibility of the experiment. The specific steps of the splitting process are as follows:

- Using the random seed S_split, 150 images are randomly selected from all samples for the test set (Test).
- Then, using the random seed S_split+1, 150 images are selected from the remaining samples for the validation set (Val).
- The remaining 550 images are allocated to the training set (Train).
- To ensure the stability of the splitting process, all samples are sorted by relative paths before splitting.

Data Augmentation

Data augmentation is an important step to enhance the model's robustness. We applied several common augmentation methods on the training set, including:

- Rotation (e.g., 90°, 180°, 270°).
- Horizontal and vertical flipping.
- Brightness adjustment.
- Random cropping and scaling.

These augmentation methods aim to expand the diversity of the dataset, helping the model better handle various input transformations. The final dataset sample distribution is shown in Table 1 of the manuscript.

Additional Notes

- The dataset splitting and augmentation processes are strictly followed as outlined above, and the same random seeds are used for each experiment.
- We plan to further enrich the dataset in future work by adding different experimental environments and wheat varieties to improve the model's performance in practical applications.