**Supplementary materials**

Table S1. Detailed information on the dataset split.

|  |  |  |
| --- | --- | --- |
|  | Training set | Test set |
| Heart | 38 | 10 |
| Experimental | 111 | 28 |
| Cherry | 112 | 29 |
| Plum | 115 | 29 |
| Total | 376 | 96 |

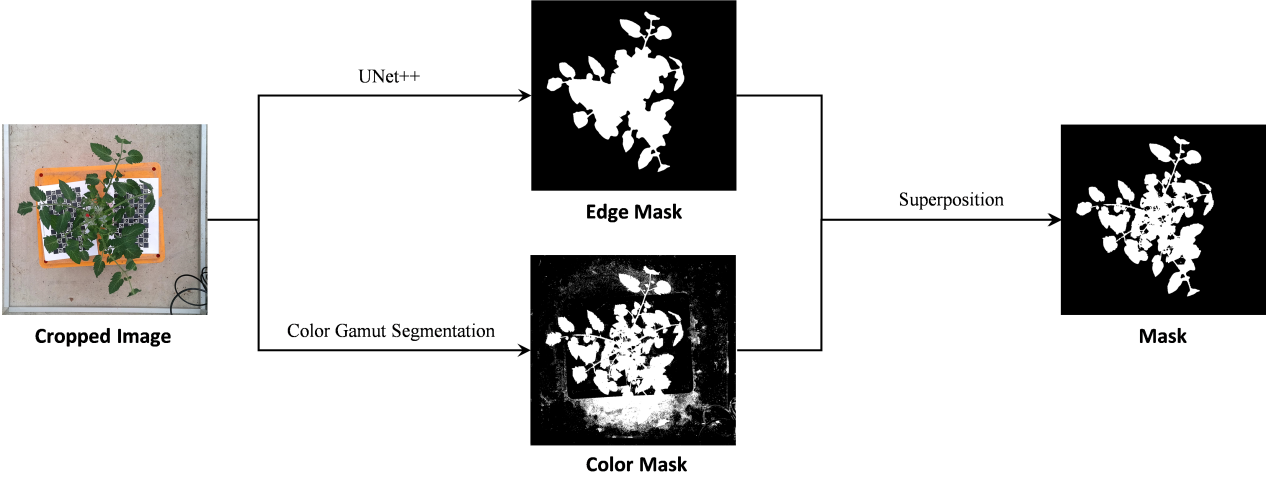


Figure S1. Workflow of mask image generation.

Table S2. The performance of UNet++ in the test set.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Class | IoU | Dice | Precision | Recall | F1 score |
| Plant | 0.96 | 0.98 | 0.98 | 0.98 | 0.98 |
| Background | 0.99 | 1.00 | 1.00 | 1.00 | 1.00 |

Table S3. The performance of YOLO v7 in the test set (all classes).

|  |  |  |  |
| --- | --- | --- | --- |
| Precision | Recall | F1 score | mAP |
| 0.83 | 0.63 | 0.72 | 0.83 |

Table S4. The performance of Random Forest. Samples with zero values for the number of red fruits are excluded from the Random Forest model.

|  |  |  |
| --- | --- | --- |
|  | R² | RMSE |
| Training set | 0.51 | 0.64 |
| Test set | 0.31 | 0.89 |

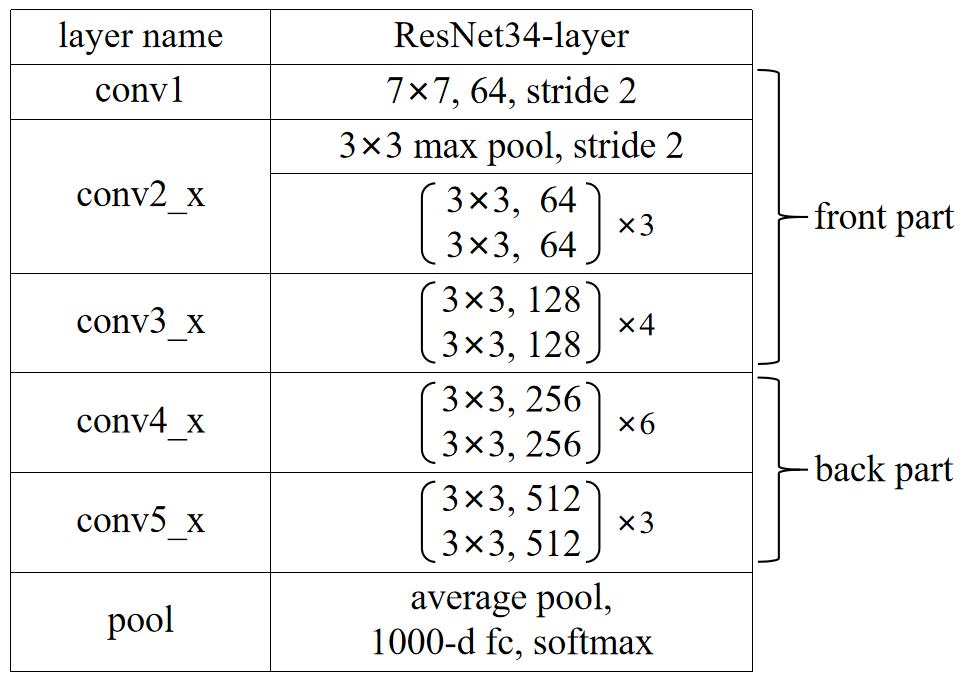


Figure S2. Structure of ResNet34. ResNet34 serves as the primary structure of TomPhenoNet. The front part of ResNet34 is used to extract features from different modalities. The back part of ResNet34 is used to predicting the values for different tasks.