

Article

Fisheye Image Detection of Trees Using Improved YOLOX for Tree Height Estimation

Jiayin Song, Yue Zhao, Wenlong Song, Hongwei Zhou *, Di Zhu, Qiqi Huang, Yiming Fan and Chao Lu

Department of Mechanical and Electrical Engineering, Northeast Forestry University, Harbin 150040, China; songjy@nefu.edu.cn (J.S.); zhaoyue@nefu.edu.cn (Y.Z.); wlSong139@126.com (W.S.); zd1998@nefu.edu.cn (D.Z.); h_qiqi@nefu.edu.cn (Q.H.); fanyiming@nefu.edu.cn (Y.F.); luhanyu@nefu.edu.cn (C.L.)

* Correspondence: easyid@163.com

Abstract: Tree height is an essential indicator in forestry research. This indicator is difficult to measure directly, as well as wind disturbance adds to the measurement difficulty. Therefore, tree height measurement has always been an issue that experts and scholars strive to improve. We propose a tree height measurement method based on tree fisheye images to improve the accuracy of tree height measurements. Our aim is to extract tree height extreme points in fisheye images by proposing an improved lightweight target detection network YOLOX-tiny. We added CBAM attention mechanism, transfer learning, and data enhancement methods to improve the recall rate, F_1 score, AP, and other indicators of YOLOX-tiny. This study improves the detection performance of YOLOX-tiny. The use of deep learning can improve measurement efficiency while ensuring measurement accuracy and stability. The results showed that the highest relative error of tree measurements was 4.06% and the average relative error was 1.62%. The analysis showed that the method performed better at all stages than in previous studies.

Keywords: tree height estimation; equidistant projection; deep learning; fisheye image



Citation: Song, J.; Zhao, Y.; Song, W.; Zhou, H.; Zhu, D.; Huang, Q.; Fan, Y.; Lu, C. Fisheye Image Detection of Trees Using Improved YOLOX for Tree Height Estimation. *Sensors* **2022**, *22*, 3636. <https://doi.org/10.3390/s22103636>

Received: 21 April 2022

Accepted: 9 May 2022

Published: 10 May 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Tree height is one of the most critical parameters in quantitative forest observation. Tree height research has important implications for urban road planning, air pollution control, and carbon neutrality. In large-scale forest stock and biomass estimation, tree height can estimate forest stock and biomass [1,2]. However, the characteristics of trees and the complex environment make direct measurements difficult. In addition, wind disturbance also increases the difficulty of measurements.

Traditional forest surveys mostly use theodolites for measurements. Theodolites can accurately obtain forest parameter factors. However, theodolites are time- and manpower-consuming. The measurement of theodolites has long survey cycles and low efficiency, so the real-time and spatial integrity of the data is difficult to keep consistent. A commonly used tool for measuring tree height is the ultrasonic rangefinder. It has the advantage of portability and real-time access to data. However, it is subject to human factors and wind speed and varies significantly from measurement to measurement. Therefore, the measurement of standing tree height remains a problem that researchers are working to improve.

Some researchers use airborne laser scanning (ALS) and terrestrial laser scanning (TLS) to measure tree heights [3–5]. However, both ALS and TLS have certain drawbacks, such as ALS is generally expensive and TLS is inconvenient to carry. Measurements using drone equipment are more costly and have poor endurance. Kedra et al. compared single-image photogrammetry (SIP) and terrestrial laser scanning (TLS). The results show that, compared to TLS, SIP can successfully apply tree-like structure feature extractions in mature forests [6]. Digital image-based measurement methods have obvious advantages in terms of economic considerations. Photogrammetry has come a long way with the development

of photography and computer vision, which has led researchers to look for new ways to measure tree heights [7,8].

The monocular vision measurement of ordinary cameras has the advantages of easy image acquisition and fewer calculation parameters required for calibration. However, normal cameras have small viewing angles and require long shooting distances when measuring large-scene objects. In 2000, Zhang proposed a tessellation grid calibration method based on pinhole cameras [9]. Before this, camera calibration often required high precision calibrators while Zhang's calibration method only required a printed checkerboard grid. After acquiring images of different directions from the checkerboard calibration plate, correspondence between the target in the 3D space and the image points on the 2D image plane can be established. After that, the internal and external parameters of the camera can be solved. However, the method is only applicable to ordinary pinhole cameras and the calibration effect is not suitable for wide-angle cameras. Scaramuzza proposed an omnidirectional camera modeling method based on the Taylor series model, which focuses on the calibration of fisheye lenses and refractive lenses within 195° [10]. The omnidirectional camera calibration method is widely used in fisheye camera calibrations because of its simple, accurate, and easy-to-use features.

Photogrammetry uses vision-based measurement methods to identify measured objects in an image. Then, it uses image processing technology to obtain coordinates of the central part of the image. The obtained coordinates are brought into the corresponding mathematical model and the measured value of the measured object can be calculated [11–13]. The extraction of extreme points in the central part of the image adopts a clustering algorithm. However, uncompressed images consume much memory which results in a long execution time for the clustering algorithm [14,15]. The image quality will be degraded after compression, which will affect measurement accuracy [16]. Researchers need to manually set the number of clusters in the clustering process based on experience [17]. In 2016, Redmon et al. first proposed the YOLO algorithm [18]. After that, the YOLO series of algorithms were widely used in agriculture, medicine, and intelligent transportation [19–21]. The YOLO series of algorithms have shown superior performance. With the continuous development of image detection algorithms, the accuracy based on deep learning has continued to improve. In 2021, Bochkovskiy and other researchers proposed YOLOv4, whose accuracy has been significantly improved compared with previous detection algorithms [22]. After YOLOv4, the YOLOX object detection network appeared and showed superior performance [23]. The YOLO series algorithm can accurately extract image feature points after training, as well as it consumes less time and does not require human experience intervention [24–27].

In this study, we propose a highly robust method for the non-contact measurements of tree height. The method uses a smartphone with a fisheye lens to capture images. The improved YOLOX algorithm is used for tree recognition and image coordinate extraction, improving recognition accuracy and efficiency.

2. Materials and Methods

2.1. Establishment of the Measurement Model

All characters and abbreviations appearing in this paper are located in Table A1 in the Appendix A. The parameters of the fisheye lens and smartphone are in Table A2 in the Appendix A. The measurement system model of this method is constructed based on the principle of the equidistant projection model. Here, $P(x_w, y_w, z_w)$ is the target point in the world coordinate system and P' is the imaging point corresponding to P in the camera coordinate system.

According to the isometric projection theorem, the projection relationship is expressed as follows:

$$r' = fw \quad (1)$$

$$w = \tan^{-1}(r/L) = \tan^{-1}[(x_w^2 + y_w^2)^{1/2}/L] \quad (2)$$

where r' is the distance from the point P' to the optical axis, f is the object square focal length of the optical system, w is the incident angle of the point P relative to the optical axis, and L is the horizontal distance between the point in the world coordinate system and the center of the fisheye lens. Due to the distortion of the fisheye lens, to ensure the uniformity of the image, the distortion coefficient λ is introduced to obtain the following:

$$r' = \lambda f w \quad (3)$$

The camera plane center point is $O_c(x_0, y_0)$, the coordinates of the P' point are (x_c, y_c) , and the coordinates of the P point are (x_w, y_w, z_w) . If the distortion coefficient components of x_c and y_c axes are λ_x and λ_y , then:

$$\begin{cases} x_c - x_0 = r' \cos \theta = \lambda_x f w \cos \theta \\ y_c - y_0 = r' \sin \theta = \lambda_y f w \sin \theta \end{cases} \quad (4)$$

$$\begin{cases} \cos \theta = x_w / (x_w^2 + y_w^2)^{1/2} \\ \sin \theta = y_w / (x_w^2 + y_w^2)^{1/2} \end{cases} \quad (5)$$

where θ is the azimuth of point P and also the azimuth of point P' in the camera coordinate system. The coordinates of the center point o in the image pixel coordinate system are (u_0, v_0) ; P' is obtained by equidistant projection P' and the relationship between the camera coordinate system and the corresponding points in the image pixel coordinate system is as follows:

$$\begin{cases} u - u_0 = m_x(x_c - x_0) = \lambda_x m_x f(x_c - x_0) \\ v - v_0 = m_y(y_c - y_0) = \lambda_y m_y f(y_c - y_0) \end{cases} \quad (6)$$

where, m_x and m_y are the amplification factors. $k_x = \lambda_x m_x f$, $k_y = \lambda_y m_y f$.

From Equations (1)–(6), the relationship between image coordinates and world coordinates is as follows:

$$\begin{cases} u = \frac{x_w k_x}{\sqrt{x_w^2 + y_w^2}} \tan^{-1} \frac{\sqrt{x_w^2 + y_w^2}}{L} + u_0 \\ v = \frac{y_w k_y}{\sqrt{x_w^2 + y_w^2}} \tan^{-1} \frac{\sqrt{x_w^2 + y_w^2}}{L} + v_0 \end{cases} \quad (7)$$

The measurement system model consists of a fisheye lens, a rangefinder, and a smartphone. The measurement system model is shown in Figure 1.

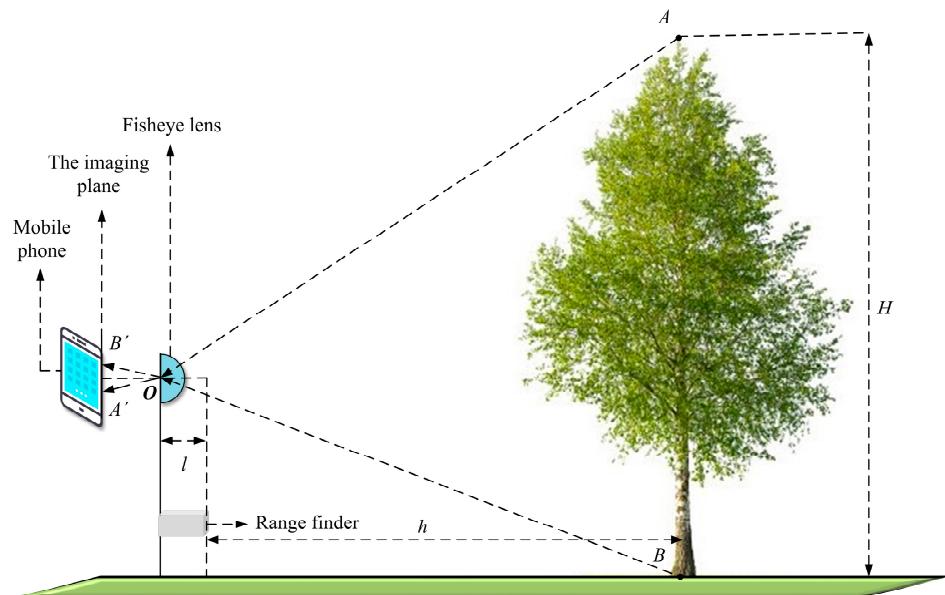


Figure 1. Measurement system model.

When using a smartphone equipped with a fisheye lens to take a picture of a single tree, $A'(u_{A'}, v_{A'})$ and $B'(u_{B'}, v_{B'})$ are the corresponding points in the image coordinate system, which are also the extreme points of the tree. The relationship between the corresponding points in the world coordinate system and the image pixel coordinate system are as follows:

$$\begin{cases} x_w = \frac{L}{\sqrt{1+\left[\frac{k_x(v-v_0)}{k_y(u-u_0)}\right]^2}} \tan \frac{(u-u_0)\sqrt{1+\left[\frac{k_x(v-v_0)}{k_y(u-u_0)}\right]^2}}{k_x} \\ y_w = \frac{k_x(v-v_0)}{k_y(u-u_0)}x \end{cases} \quad (8)$$

The coordinates of the center point of the image coordinate system are $o(u_0, v_0)$. k_x and k_y are the distortion coefficients of the fisheye image, which can be obtained by the camera calibration method. L is the horizontal distance, which the following formula can obtain:

$$L = h + l \quad (9)$$

where h is the horizontal distance in the world coordinate system and l is the virtual imaging distance of the fisheye lens. Through the transformation relationship between coordinate systems, the following formula can be obtained:

$$\begin{cases} x_w = \frac{L}{\sqrt{1+\left[\frac{k_x(v-v_0)}{k_y(u-u_0)}\right]^2}} \tan \frac{(u-u_0)\sqrt{1+\left[\frac{k_x(v-v_0)}{k_y(u-u_0)}\right]^2}}{k_x} \\ y_w = \frac{k_x(v-v_0)L}{k_y(u-u_0)\sqrt{1+\left[\frac{k_x(v-v_0)}{k_y(u-u_0)}\right]^2}} \tan \frac{(u-u_0)\sqrt{1+\left[\frac{k_x(v-v_0)}{k_y(u-u_0)}\right]^2}}{k_x} \end{cases} \quad (10)$$

According to Equation (10), H is the result obtained by the measurement system model [28].

$$H = [(x_A - x_B)^2 + (y_A - y_B)^2]^{1/2} \quad (11)$$

In Equation (11), H is the final calculated tree height value; the extreme points $A'(u_{A'}, v_{A'})$ and $B'(u_{B'}, v_{B'})$ of the tree are the parameters needed to calculate the tree height.

Tree extrema are defined as the highest and lowest points of a tree. The improved YOLOX-tiny object detection network can detect the complete tree and extract tree extreme points. After that, the extracted extreme points are brought into the tree height estimation model to calculate the tree's height. The general flow chart for estimating tree height is shown in Figure 2.

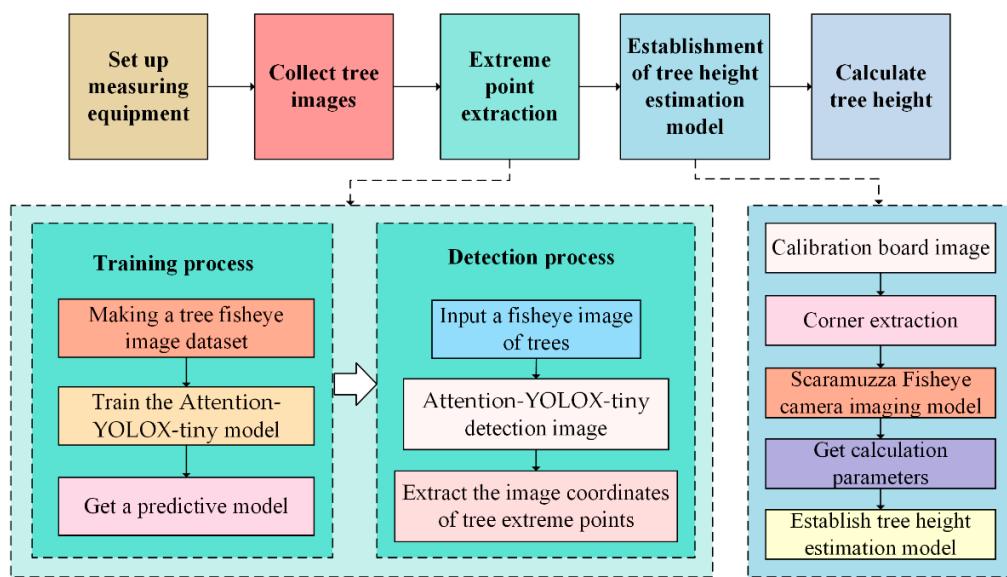


Figure 2. Tree height calculation process.

The procedure for calculating tree height is as follows:

- Set up measuring equipment. A smartphone with a fisheye lens is required to set up the measuring equipment.
- Acquire images. After training is complete, only one image of the tree under test needs to be collected.
- Extract extreme points. Deep learning methods can perform this step quickly and accurately.
- Build a tree height calculation model. This step only needs to be done once during the initial calculation.
- Calculate tree height. Obtain results and perform error analysis.

2.2. Improved Target Detection Network

YOLOX is similar to the previous YOLO version. The whole YOLOX can be divided into the following parts: CSPDarknet is the backbone feature extraction network of YOLOX. The input image is extracted in CSPDarknet and the extracted features are the feature layer, which is the feature set of the input image. FPN is an enhanced feature extraction network of YOLOX. The feature extraction module is performed using the obtained effective feature layers. YOLOX not only upsamples the fused features but also downsamples the fused features. YOLO Head is a classifier of YOLOX with three enhanced effective feature layers obtained by CSPDarknet and FPN. Each feature layer has a width, height, and number of channels. YOLOX uses the Focus network structure, which is used in YOLOV5. In a picture, every other pixel takes a value to get four independent feature layers and then these four separate feature layers are stacked. First, the input image is subjected to shallow feature extraction. Then, the three feature layers are outputted to the feature fusion part for deep feature extraction.

The attention mechanism refers to the panorama of the image that the human vision can focus on a certain local area. The attention mechanism is also used in the research of deep learning. The idea is to use new weights to highlight key points in the image data and train the network. The model identifies the location of the target of interest in the dataset. CBAM (Convolutional Block Attention Module) is a lightweight attention module [29].

CBAM first learns the weight distribution from the relevant features. Then, it feeds the weights back to the features to enhance the network feature recognition ability. The convolutional layer plays a crucial role in the process of feature extraction. The number of channels in each convolution layer is only related to the number of convolution kernels. The feature map is the result of the convolution operation of the input image. However, the convolution layer contains many convolution kernels and the generated feature map

will also have a corresponding number of channels. The existence of the attention model plays a role in channel filtering.

In this study, embedding the CBAM module keeps the original YOLOX-tiny structure (Attention-YOLOX-tiny). The network structure of Attention-YOLOX-tiny is shown in Figure 3.

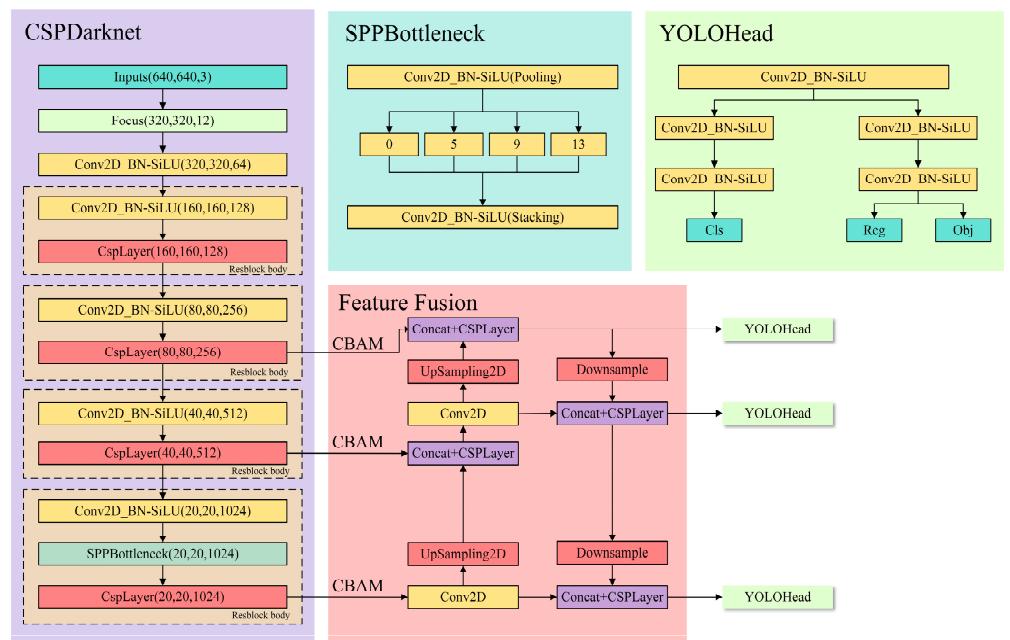


Figure 3. Attention-YOLOX-tiny network structure.

To obtain better detection results, the transfer learning method is used to load the pre-trained model [30]. The learning model needs to learn related source tasks on the source domain and then transfer the knowledge to the target task on the target domain to improve the model's performance on specific tasks. Given the source domain (D_S) and the source task (T_S), the target domain (D_T) and target task (T_T), the knowledge acquired, and T_S help the model solve the prediction function (f_T) of T_T on D_T . The transfer learning process is shown in Figure 4.

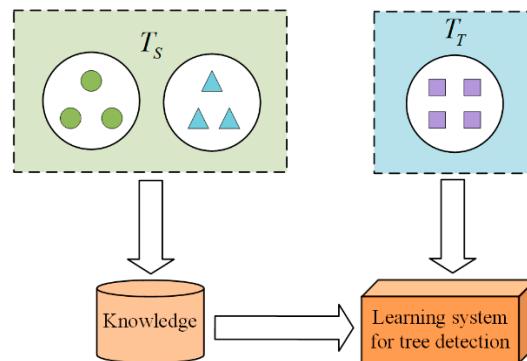


Figure 4. Transfer learning.

During the training process, mosaic data augmentation is used to augment the dataset. Mosaic data enhancement refers to reading four pictures at a time, flipping, scaling, and changing the color gamut of the four pictures, respectively. Then, it positions according to the positions of the four directions and combines the pictures and frames. The mosaic data enhancement method is shown in Figure 5.

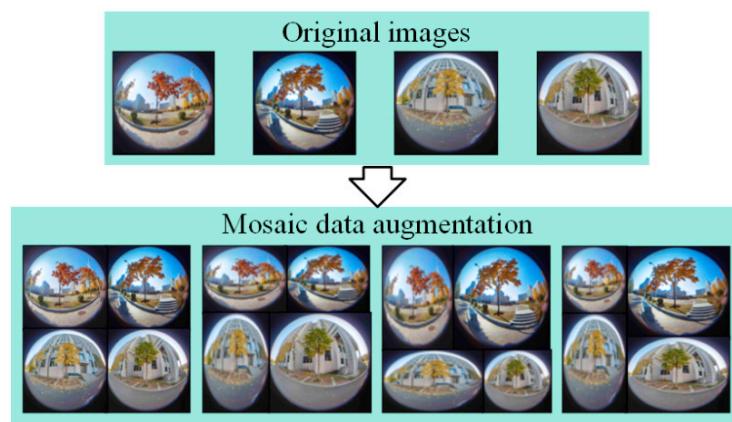


Figure 5. Mosaic data augmentation.

3. Experimental Results and Analysis

3.1. Validation of Fisheye Lens Measurement Model

After obtaining the required parameters of the tree height measurement model, the tree height measurement model is used to measure the distance between the corner points of the black and white chessboard. The measurement model is verified by comparing the actual distance between the corner points.

The optical centroid is found using the Scaramuzza model and the fisheye image is processed by directly calling the matlab2018b fisheye lens calibration toolbox. The extraction of the checkerboard and checkerboard corners is shown in Figure 6.

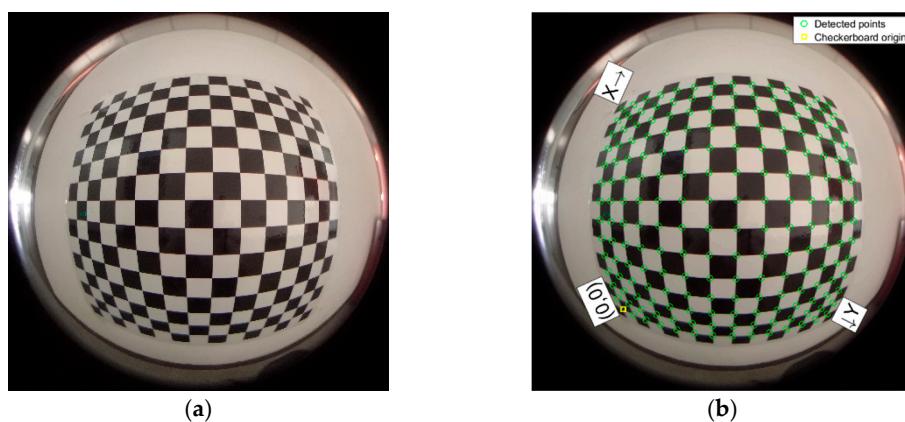


Figure 6. Checkerboard and corner extraction. (a) Fisheye image; (b) Corner extraction.

The following steps were performed: camera calibration, calculation of the distortion coefficient corresponding to all corners in each chessboard, and the average value was taken for subsequent calculations. Five sets of chessboard diagrams with different distances were taken to verify the accuracy of the measurement model, as shown in Figure 7.

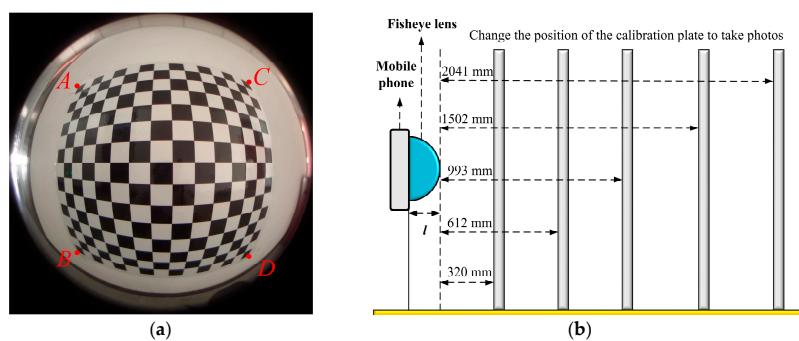


Figure 7. Corner point position selection and calibration plate position selection. (a) Corner point position, A , B , C , and D are corner points; (b) Five positions of the calibration plate.

Three sets of distances, AB , AC , and AD , were taken on the chessboard, the error was analyzed, and the accuracy of the measurement model was verified. Table 1 shows the calculation results of the distance between AB , AC , and AD .

Table 1. AB , AC , AD calculation results.

The analysis and calculation results show that the average relative error is 0.823%. The measurement error of this measurement model is low and can be applied to the measurement of tree height.

3.2. Tree Detection and Extreme Point Extraction

Before taking pictures, 178 randomly selected trees from the Northeast Forestry University were marked. To improve the robustness of the model and fully consider the effect of light in the experiments, a smartphone equipped with a fisheye lens was used to capture fisheye images on sunny and cloudy days. We acquired 1035 photos (including 537 on sunny days and 498 on cloudy days). The image acquisition time covers the whole day. The dataset contains different light intensities which ensures the adaptability of the method to different light intensities and improves the robustness of the prediction model. Figure 8 shows annotation results of the fisheye images under different weather conditions.

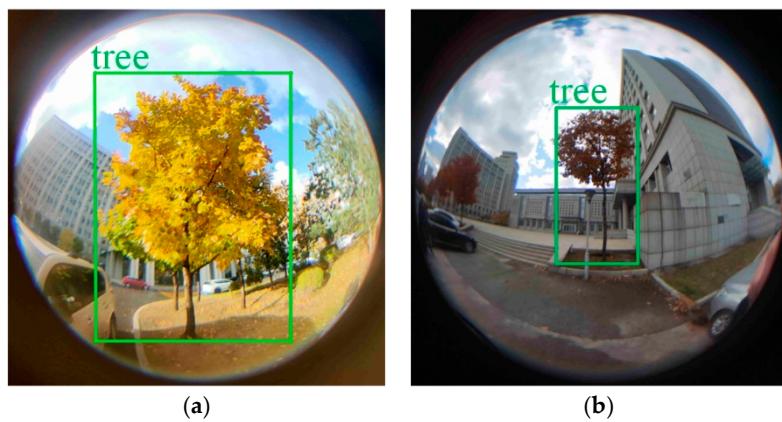


Figure 8. Fisheye image annotation under different weather conditions. (a) Sunny images annotation; (b) Cloudy images annotation.

To explore the network structure of YOLOX with the best detection effect, YOLOX-s, YOLOX-tiny, and Attention-YOLOX-tiny are tested. The precision (P), recall (R), F_1 score (F_1), and average precision (AP) are used to evaluate the target detection model. P is for the prediction result and it is the proportion of correctly predicted positive samples to all predicted samples. R is for the original sample and it is the proportion of correctly predicted positive samples out of all positive samples. F_1 is an indicator and a trade-off of P and R. AP is the area under the P-R curve. AP can measure the trained model on a single tree prediction. These evaluation indicators are defined as the following formulas:

$$P = \frac{TP}{TP + FP} \quad (12)$$

$$R = \frac{TP}{TP + FN} \quad (13)$$

$$F_1\text{score} = \frac{2P \times R}{P + R} \quad (14)$$

$$AP = \int_0^1 P(R)dR \quad (15)$$

The acquired images were used to make a fisheye image dataset. To increase the training efficiency, Docsmall (an image compression website) was used to compress the images before training. The compressed images were divided into a training set and validation set. The ratio of the training set and validation set was 9:1. The processor used for training was Intel Core i7-10700K, 3.80 GHZ processor, 32 GB memory, 10 GB NVIDIA RTX 3080 GPU. The training parameters are set as shown in Table 2. The evaluation results of each network are shown in Table 3.

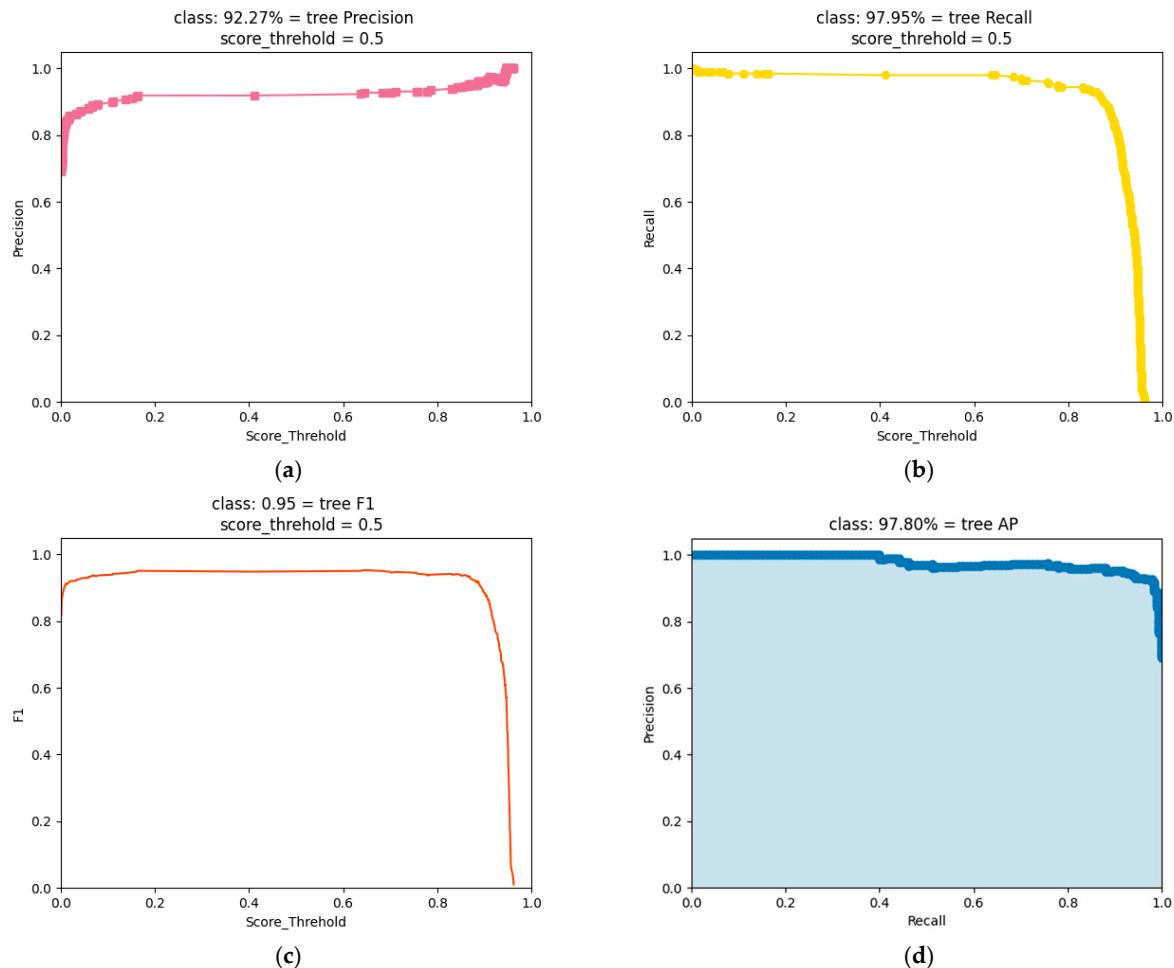
Table 2. Training parameters.

Parameters	Value
Input size	640×640
Output size	640×640
Learning rate	adaptive
Batch size	8
Epoch	300

Table 3. Detection evaluation of different networks.

Model	Epoch	P (%)	R (%)	F ₁	AP (%)
YOLOX-s	300	92.57	95.90	0.94	96.27
YOLOX-tiny	300	93.03	95.90	0.94	97.26
Attention-YOLOX-tiny	300	92.27	97.95	0.95	97.80

As shown in Table 3, on the fisheye image dataset of trees, the P of Attention-YOLOX-tiny is 92.27%, R is 97.95%, F₁ is 0.95, and AP is 97.80%. Most of the performance metrics of Attention-YOLOX-tiny, including R, F₁, and AP, are better than YOLOX-s and YOLOX-tiny. The evaluation process of the Attention-YOLOX-tiny detection model is shown in Figure 9.

**Figure 9.** Evaluation results, which include P, R, F₁, and AP. (a) P; (b) R; (c) F₁; (d) AP.

From the detection indicators in Figure 9, it can be concluded that the prediction results of trees can meet the requirements of accurate detection of trees. The LOSS function curve of the training process is shown in Figure 10.

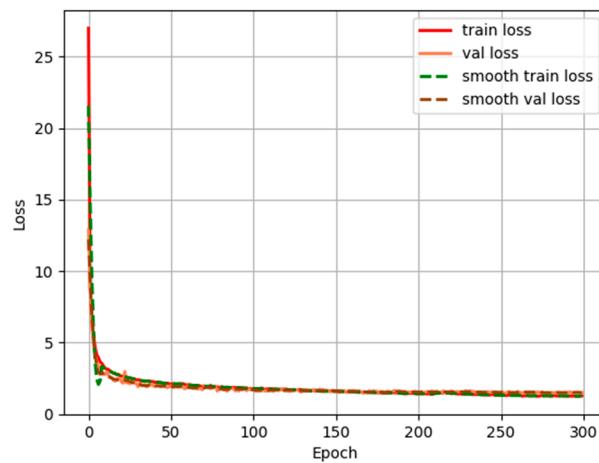


Figure 10. LOSS function change curve.

Figure 10 shows the loss variation curve of Attention-YOLOX-tiny, where the horizontal and vertical axes represent training epochs and loss values. With the increasing number of training iterations, the loss value on the training set, the loss value on the validation set, the smooth loss value on the training set, and the smooth loss value on the validation set of Attention-YOLOX-tiny all decrease rapidly at first, and then gradually decrease. The loss curve of Attention-YOLOX-tiny gradually converges around 2.0 after about 150 iterations. The loss curve has converged which indicates that the predicted output is credible. The trained model has learned the characteristics of the tree under the fisheye distortion and can extract the extreme points.

Figure 11 shows the detection result of the tree and the extraction result of the extreme points of the tree. Through the above experimental analysis, it can be concluded that Attention-YOLOX-tiny can accurately detect the target object in the picture. By extracting the coordinates of the detection frame in the picture, the coordinates of the extreme point $A'(u_{A'}, v_{A'})$ and the extreme point $B'(u_{B'}, v_{B'})$ in the model can be obtained.



Figure 11. Tree detection and tree extreme point extraction. (a) Fisheye image of detection tree; (b) Extract extreme points.

3.3. Tree Height Calculation

The coordinates of $A'(u_{A'}, v_{A'})$ and $B'(u_{B'}, v_{B'})$ are the coordinates of the midpoints of the upper and lower frame lines in the image. The fisheye lens measurement model is taken to obtain the predicted tree height. In this experiment, the average value of the ten times measured by the theodolite was taken as the actual value. We selected 83 trees as validation data and for contrastive measurements with Transponder T3. The tree height measurement results are shown in Figure 12; Figure 12a is the comparison of the relative errors of the fisheye model and Transponder T3 and Figure 12b is the comparison of their measurement values.

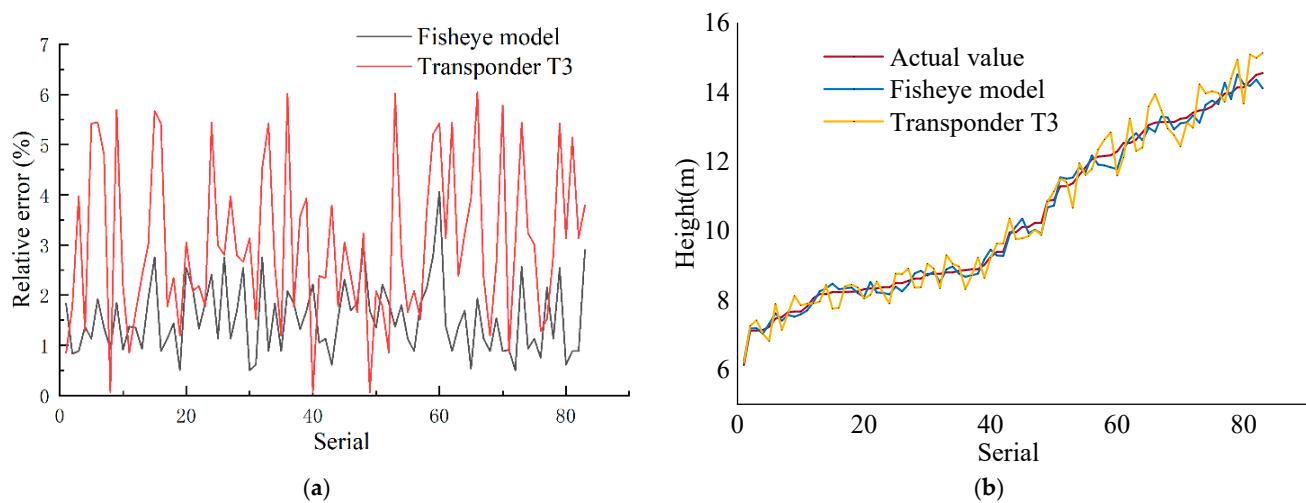


Figure 12. Measurement results. (a) Measurement Error; (b) Measured value.

The experimental results show that the average relative error of the method in this paper is 1.62% and the average relative error of Transponder T3 is 3.23%. Through comparison, it can be found that the average error of this method is significantly smaller than Transponder T3. The calculation result of this method is more stable than Transponder T3.

3.4. Wind Interference Experiment

The measurement environment in practical applications is variable. To verify the accuracy of this method under windy measurement conditions, a windy day was selected. The wind level was 5~6 (taken from China Weather Network). Transponder T3 does not work correctly in this condition. The fisheye images of 30 trees were obtained and calculated. The experiment shows that under the conditions of wind measurement, the average error of this method is 2.31% and Transponder T3 has completely failed. The calculation result of this method is shown in Figure 13. The shaded part in the Figure 13 is the absolute error. The practicality of this method under the influence of wind is better than that of Transponder T3.

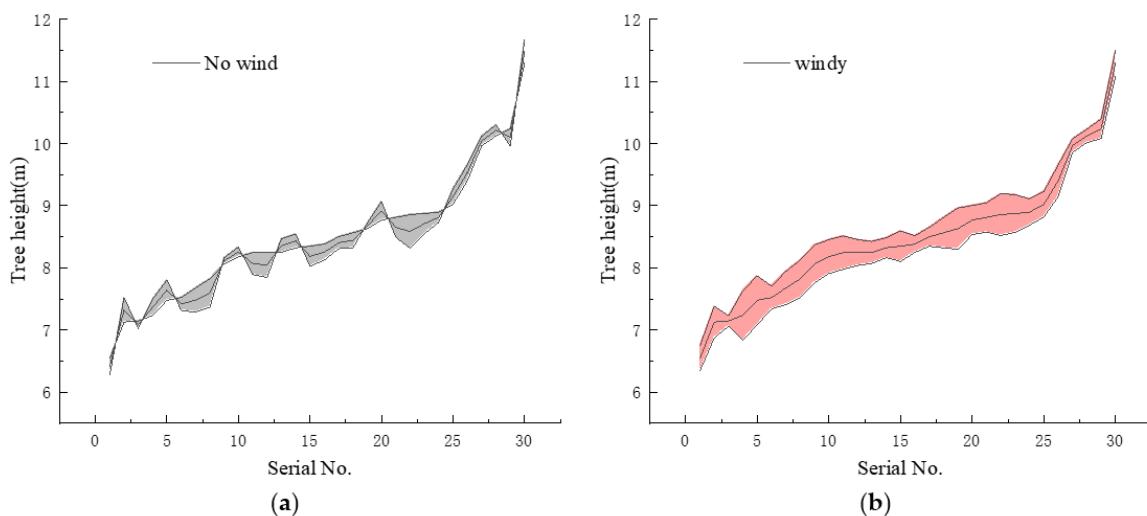


Figure 13. Errors under different measurement conditions. (a) The wind is less than level 3; (b) The wind is at level 5~6.

When the wind reaches level 5~6, the measurement effect of the method in this paper is affected because the wind changes the shape of the tree and affects the extreme point

coordinates of the tree extracted by Attention-YOLOX-tiny. This eventually leads to an increase in measurement error.

4. Conclusions

Compared with the ultrasonic rangefinder to measure tree height, the relative error of the ultrasonic rangefinder was the highest at 6.04%, the lowest was 0.34%, and the average relative error was 3.23%. The highest relative error of the method calculated in this paper is 4.06%, the lowest relative error is 0.5%, and the average relative error is 1.62%. In tree detection, Attention-YOLOX-tiny can accurately and quickly extract the extreme points of trees. Overall, the average relative error of the method in this paper is low, which is better than the ultrasonic rangefinder in measurement accuracy. The method has the advantages of stable measurement, compact structure, and easy portability.

Experiments were carried out to analyze the errors under different measurement conditions. The average relative error of the method in this paper is 2.31% under the condition of level 5–6 wind. Compared with the no-wind condition, the relative error calculated by this method increases slightly under the gale conditions. However, it can still complete the measurement task and maintain good accuracy.

As an important indicator for measuring forest carbon storage, tree height has always been a hotspot in forest research. This study obtains Attention-YOLOX-tiny by improving the target detection network and proposes a new method for measuring tree height based on Attention-YOLOX-tiny. Consisting of a mobile phone and a matching fisheye lens, the measurement device will continue to improve with the rapid development of electronics and manufacturing capabilities. The proposal of more accurate and lightweight detection networks in computer vision can extract the extreme points of trees more quickly and accurately. In future research, tree extremum points can be extracted by faster and more accurate object detection and segmentation networks. The disadvantage of this research is that it is difficult to obtain 3D information about trees only through 2D images; 3D reconstruction of trees through images is the main break-through direction in the future.

Author Contributions: Conceptualization, J.S. and Y.Z.; methodology, J.S.; software, H.Z.; validation, W.S.; formal analysis, J.S.; investigation, J.S.; resources, Y.Z.; data curation, D.Z.; writing—original draft preparation, J.S.; writing—review and editing, J.S.; visualization, Q.H.; supervision, C.L.; project administration, Y.F.; funding acquisition, D.Z. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by “The Fundamental Research Funds for the Central Universities”, grant number: 2572017CB13, and by “Heilongjiang Provincial Natural Science Foundation of China”, grant number: YQ2020C018.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Not applicable.

Acknowledgments: Thanks for the experimental environment and experimental equipment provided by the Computer Vision Laboratory of Northeast Forestry University.

Conflicts of Interest: The authors declare no conflict of interest.

Appendix A

Table A1. Symbols and abbreviations which appear in the text.

Symbol or Abbreviation	Explanation
P	The target point in the world coordinate system.
P'	The imaging point corresponding to P in the camera coordinate system.
r'	The distance from the point P' to the optical axis.
f	The object square focal length of the optical system.
w	The incident angle of the point P relative to the optical axis.
L	The horizontal distance between the point in the world coordinate system and the center of the fisheye lens.
λ	The distortion coefficient.
θ	The azimuth of point P and the azimuth of point P' in the camera coordinate system.
m_x and m_y	The amplification factors.
k_x and k_y	The distortion coefficients of the fisheye image.
h	The horizontal distance in the world coordinate system.
H	The result was obtained by the measurement system model.
CBMA	Convolutional Block Attention Module.
D_S	Source domain.
T_S	Source task.
D_T	Target domain.
T_T	Target task.
f_T	Target function.

Table A2. Measuring equipment parameters.

Fisheye Lens			Smartphone		
Attributes	Value	Unit	Attributes	Value	Unit
Thread diameter	17	mm	Size	148.9 × 71.1 × 8.5	mm
Angle	180		Pixel	50	million
Weight	36	g	Weight	175	g
Resolution	4096 × 4096	dpi	Photo resolution	8192 × 6144	dpi

References

1. Huang, Y.D.; Li, M.Z.; Ren, S.Q.; Wang, M.J.; Cui, P.Y.J.B. Impacts of tree-planting pattern and trunk height on the airflow and pollutant dispersion inside a street canyon. *Build. Environ.* **2019**, *165*, 106385. [[CrossRef](#)]
2. Calvo-Alvarado, J.C.; McDowell, N.G.; Waring, R.H.; Physiology, R.H.J.T. Allometric relationships predicting foliar biomass and leaf area:sapwood area ratio from tree height in five Costa Rican rain forest species. *Tree Physiol.* **2008**, *28*, 1601–1608. [[CrossRef](#)] [[PubMed](#)]
3. Wang, Y.; Lehtomäki, M.; Liang, X.; Pyörälä, J.; Kukko, A.; Jaakkola, A.; Liu, J.; Feng, Z.; Chen, R.; Hyyppä, J. Is field-measured tree height as reliable as believed—A comparison study of tree height estimates from field measurement, airborne laser scanning and terrestrial laser scanning in a boreal forest. *ISPRS J. Photogramm. Remote Sens.* **2019**, *147*, 132–145. [[CrossRef](#)]
4. Parent, J.R.; Volin, J.C. Assessing species-level biases in tree heights estimated from terrain-optimized leaf-off airborne laser scanner (ALS) data. *Int. J. Remote Sens.* **2015**, *36*, 2697–2712. [[CrossRef](#)]
5. Calders, K.; Adams, J.; Armston, J.; Bartholomeus, H.; Bauwens, S.; Bentley, L.P.; Chave, J.; Danson, F.M.; Demol, M.; Disney, M.; et al. Terrestrial laser scanning in forest ecology: Expanding the horizon. *Remote Sens. Environ.* **2020**, *251*, 112102. [[CrossRef](#)]
6. Kędra, K.; Barbeito, I.; Dassot, M.; Vallet, P.; Gazda, A.J. Single-image photogrammetry for deriving tree architectural traits in mature forest stands: A comparison with terrestrial laser scanning. *Ann. For. Sci.* **2019**, *76*, 5. [[CrossRef](#)]
7. Eliopoulos, N.J.; Shen, Y.; Luong, N.M.; Vaastav, A.; Zhang, Y.; Shao, G.; Keith, W.; Lu, Y.-H. Rapid Tree Diameter Computation with Terrestrial Stereoscopic Photogrammetry. *J. For.* **2020**, *118*, 355–361. [[CrossRef](#)]
8. Zagalikis, G.; Cameron, A.D.; Miller, D.R. The application of digital photogrammetry and image analysis techniques to derive tree and stand characteristics. *Can. J. For. Res.* **2005**, *35*, 1224–1237. [[CrossRef](#)]
9. Zhang, Z. A Flexible New Technique for Camera Calibration. *IEEE Trans. Pattern Anal. Mach. Intell.* **2000**, *22*, 1330–1334. [[CrossRef](#)]

10. Scaramuzza, D.; Martinelli, A.; Siegwart, R. A Toolbox for Easily Calibrating Omnidirectional Cameras. In Proceedings of the IEEE/RSJ International Conference on Intelligent Robots & Systems, Benjing, China, 9 February 2009.
11. Shen, F.; Qin, F.; Zhang, Z.; Xu, D.; Wu, W. Automated Pose Measurement Method Based on Multivision and Sensor Collaboration for Slice Micro Device. *IEEE Trans. Ind. Electron.* **2020**, *68*, 498. [[CrossRef](#)]
12. Zhou, Y.; Li, Q.; Wu, Y.; Ma, Y.; Wang, C.J. Trinocular vision and spatial prior based method for ground clearance measurement of transmission lines. *Appl. Opt.* **2021**, *60*, 2422–2433. [[CrossRef](#)] [[PubMed](#)]
13. Mao, J.; Huang, W.; Sheng, W.J. Target distance measurement method using monocular vision. *IET Image Process.* **2020**, *14*, 3181–3187.
14. Isa, N.A.M.; Mat, N.A.; Salamah, S.A.; Samy, A.; Ngah, U.K.; Kaithum, U.J. Adaptive Fuzzy Moving K-means Clustering Algorithm for Image Segmentation. *IEEE Trans. Consum. Electron.* **2009**, *55*, 2145–2153.
15. Jaisakthi, S.M.; Murugaiyan, S.; Mirunalini, P.; Aravindan, C. Automated skin lesion segmentation of dermoscopic images using GrabCut and k-means algorithms. *IET Comput. Vis.* **2018**, *12*, 1088–1095. [[CrossRef](#)]
16. Liu, Z.-y.; Ding, F.; Xu, Y.; Han, X. Background dominant colors extraction method based on color image quick fuzzy c-means clustering algorithm. *Def. Technol.* **2020**, *17*, 1782–1790. [[CrossRef](#)]
17. Dhal, K.G.; Das, A.; Ray, S.; Gálvez, J. Randomly Attracted Rough Firefly Algorithm for histogram based fuzzy image clustering. *Knowl.-Based Syst.* **2021**, *216*, 106814. [[CrossRef](#)]
18. Redmon, J.; Divvala, S.; Girshick, R.; Farhadi, A.J.I. You Only Look Once: Unified, Real-Time Object Detection. *arXiv* **2016**, arXiv:1506.02640.
19. Xu, H.; Guo, M.; Nedjah, N.; Zhang, J.; Li, P. Vehicle and Pedestrian Detection Algorithm Based on Lightweight YOLOv3-Promote and Semi-Precision Acceleration. *IEEE Trans. Intell. Transp. Syst.* **2022**, *1*–12. [[CrossRef](#)]
20. Wang, D.; He, D. Channel pruned YOLO V5s-based deep learning approach for rapid and accurate apple fruitlet detection before fruit thinning. *Biosyst. Eng.* **2021**, *210*, 271–281. [[CrossRef](#)]
21. Mohamadipanah, H.; Kearse, L.D.; Witt, A.; Wise, B.; Pugh, C. Can Deep Learning Algorithms Help Identify Surgical Workflow and Techniques? *J. Surg. Res.* **2021**, *268*, 318–325. [[CrossRef](#)]
22. Bochkovskiy, A.; Wang, C.Y.; Liao, H.Y.M. YOLOv4: Optimal Speed and Accuracy of Object Detection. *arXiv* **2020**, arXiv:2004.10934.
23. Ge, Z.; Liu, S.; Wang, F.; Li, Z.; Sun, J. YOLOX: Exceeding YOLO Series in 2021. *arXiv* **2021**, arXiv:2107.08430.
24. Wang, N.; Ri, K.; Liu, H.; Zhao, X. Structural Displacement Monitoring Using Smartphone Camera and Digital Image Correlation. *IEEE Sens. J.* **2018**, *18*, 4664–4672. [[CrossRef](#)]
25. Yu, L.; Tao, R.; Lubineau, G. Accurate 3D Shape, Displacement and Deformation Measurement Using a Smartphone. *Sensors* **2019**, *19*, 719. [[CrossRef](#)]
26. Yu, L.; Lubineau, G. A smartphone camera and built-in gyroscope based application for non-contact yet accurate off-axis structural displacement measurements. *Measurement* **2020**, *167*, 108449. [[CrossRef](#)]
27. Groote, F.D.; Vandevyvere, S.; Vanhevel, F.; Xivry, J.J.O.D.J.G. Validation of a smartphone embedded inertial measurement unit for measuring postural stability in older adults. *Gait. Posture* **2021**, *84*, 17–23. [[CrossRef](#)] [[PubMed](#)]
28. Song, J.; Zhao, Y.; Chi, Z.; Ma, Q.; Yin, T.; Zhang, X. Improved FCM algorithm for fisheye image cluster analysis for tree height calculation. *Math. Biosci. Eng.* **2021**, *18*, 7806–7836. [[CrossRef](#)] [[PubMed](#)]
29. Woo, S.; Park, J.; Lee, J.Y.; Kweon, I.S.J.S. Cham. CBAM: Convolutional Block Attention Module. *arXiv* **2018**, arXiv:1807.06521.
30. Opbroek, A.V.; Ikram, M.A.; Vernooij, M.W.; de Bruijne, M. Transfer learning improves supervised image segmentation across imaging protocols. *IEEE Trans. Med. Imaging* **2015**, *34*, 1018. [[CrossRef](#)]