Intelligent Localization Sampling System Based on Deep Learning and Image Processing Technology

Shengxian Yi, Zhongjiong Yang *, Liqiang Zhou, Shaoxin Zou and Huangxin Xie

Abstract: In this paper, deep learning and image processing technologies are combined, and an automatic sampling robot is proposed that can completely replace the manual method in the three-dimensional space when used for the autonomous location of sampling points. It can also achieve good localization accuracy, which solves the problems of the high labor intensity, low efficiency, and poor scientific accuracy of the manual sampling of mineral powder. To improve localization accuracy and eliminate non-linear image distortion due to wide-angle lenses, distortion correction was applied to the captured images. We solved the problem of low detection accuracy in some scenes of Single Shot MultiBox Detector (SSD) through data augmentation. A visual localization model has been established, and the image coordinates of the sampling point have been determined through color screening, image segmentation, and connected body feature screening, while coordinate conversion has been performed to complete the spatial localization of the sampling point, guiding the robot in performing accurate sampling. Field experiments were conducted to validate the intelligent sampling robot, which showed that the maximum visual positioning error of the robot is 36 mm in the x-direction and 24 mm in the y-direction, both of which meet the error range of less than or equal to 50 mm, and could meet the technical standards and requirements of industrial sampling localization accuracy.

Keywords: visual localization; deep learning; image processing; SSD; intelligent sampling robot; area sensing; automatic sampling of mineral powder

1. Introduction

In the fiercely competitive environment of the metallurgical industry, the efficiency of ore powder quality inspection is directly related to the costs and benefits of the enterprise. As shown in Figure 1, enterprises must randomly sample several areas of the transport carriage, and manual inspections, with high labor intensity and low operating efficiency, will directly expose the inspector to the risk of heavy metal pollution. If automated robots are used to sample, it can not only reduce labor costs and improve efficiency, but also avoid the risk of corruption; therefore, automation and intelligence are obviously important directions for the future development of the mineral and industrial fields [1–3]. However, many of the currently used mine powder sampling machines have shortcomings, such as their small sampling range, insufficient intelligent working ability, long sampling and localization time, and insufficient sampling localization accuracy. Traditional conventional sampling techniques can no longer meet the needs of modernization [4–8].

In recent years, target detection algorithms based on deep learning have been widely used in various detection scenarios [9], mainly because of their strong comprehensiveness, activity, and ability to detect and recognize multiple types of objects at the same time [10–13]. They can be roughly divided into two categories—two-stage algorithms based on the candidate area, such as the R-CNN series, fast R-CNN, faster R-CNN, etc. [14–16],...
and one-stage algorithms based on the regression method, such as YOLO (You Only Look Once) series and SSD [17–21]. Two-stage algorithms have high accuracy but slow speed, and are not suitable for the real-time requirements of ore powder target detection. One-stage algorithms have higher speed, better real-time performance in target detection, and their accuracy can reach the level of FASTER R-CNN [22]. Therefore, the one-stage algorithm is more suitable for the target detection of ore powder than the two-stage algorithm.

![Figure 1. Manual random sampling. (a) Sampling area division. (b) Manual sampling.](image)

The SSD algorithm and YOLO algorithm, as typical representatives of one-stage algorithms, are widely used in the engineering field [23]. However, the YOLO algorithm also has certain limitations. The YOLO algorithm has a weak generalization ability for object aspect ratio, and its detection accuracy decreases when the aspect ratio of a new class of objects appears [24–26]. SSD has two advantages, namely, real-time processing and high accuracy, especially when targeting objects in different size scales, meaning a certain level of accuracy can be guaranteed [27–31]. Given that the size of the ore hauling vehicle to be detected is generally variable and uncertain, the SSD algorithm is comparatively more suitable for the multi-scale detection of ore hauling vehicles. Thus, given the detection task of this study, the SSD algorithm was selected as the detection model for the sampling area.

However, SSD has the limitation that small objects are not detected well. This is not only a problem for SSD, but also for most object detection algorithms. To solve this problem, various attempts have been made at replacing VGGNet with ResNet [32,33] or feature fusion [34]. Fu et al. proposed another version of SSD called Deconvolutional SSD (DSSD), which used ResNet instead of VGGNet and achieved better performance compared to the conventional SSD; however, DSSD improved accuracy at the cost of speed [35]. Jeong et al. improved the method of feature fusion, so as to make full use of the features of each output layer [36]. Li et al. proposed the feature fusion Single-Shot MultiBox Detector (FSSD) model, which obtained more details of the output feature layers through feature fusion and down-sampling, so as to improve the detection accuracy of the model [37]. Although the improved SSD algorithm can increase the detection accuracy, the automatic detection of mineral powder involves not only detecting the area of the carriage, but also detecting and locating the 3 × 3-pixel locating light spot in the image. The current SSD algorithm cannot achieve such accuracy, even though it is improved temporarily, and it cannot achieve direct detection, so it needs to be integrated with other image technologies to enable automated localization and sampling detection.

Currently, image processing techniques, such as color screening and image segmentation, are widely used and studied. In terms of HSV, Qussay Al-Jubouri et al. proposed a method for classifying individual zebrafish based on statistical texture and hue/saturation/value (HSV) color features. Experimental comparisons show that the effectiveness of this method is significantly improved compared with previous individual classification methods [38]. Tomáš Harashty et al. proposed a traffic sign detector based on the HSV color model, and the success rate of the detector exceeded 95% when using the HSV color model [39]. Dede Wand et al. applied a fusion of HSV and HSI methods to detect the wilting condition of roses, and concluded that this method was the fastest [40].
K. Cantrell et al. used the hue parameters of hue, saturation, and value color space as quantitative analytical parameters for bitonal optical sensors. Studies have shown that the H value can maintain excellent accuracy even under changes of indicator concentration, film thickness, and detector spectrum, showing its great prospects for application in various sensing applications [41]. In addition, in terms of image segmentation, Wang et al. introduced a new automatic Region-based Image Segmentation Algorithm based on k-means clustering (RISA), specifically designed for remote sensing applications. This method was shown to have better flexibility and accuracy through case studies [42]. Silvia Ojeda et al. introduced a new algorithm to perform image segmentation; the foundations of this algorithm are random field theory and its robustness when used for spatial autoregressive processes. Experimental results with real images have verified how the algorithm works in practice [43]. Siddharth Singh Chouhan et al. investigated the existing application of Soft Computing techniques in image segmentation, and their research showed that the use of Soft Computing methods can deal with uncertainty and inaccuracy, thus enabling intelligent systems to gain flexibility and adaptability [44]. According to the current research, image processing technology is widely used in various fields related to images owing to its superiority in real time, its adaptability, and its high accuracy. In this study, the SSD algorithm based on deep learning is combined with image processing technologies such as color screening and image segmentation to achieve accurate localization and coordinate transformation of sampling points, so as to guide the robot to conduct automatic detection and meet the requirements of the error range of localization. Compared with the traditional conventional sampling technology, it has the advantages of shorter time-consuming sampling and positioning, stronger intelligent working ability, and larger sampling range.

The main contributions of this study are as follows:

(1) Using the checkerboard calibration board and the MATLAB toolbox, the camera’s internal parameters and the distortion parameters have been determined. Then, the acquired images were distortion-corrected by OpenCV (Open Source Computer Vision Library) to improve the accuracy of sampling area detection and visual localization. The use of data augmentation has solved the problem of the lower availability of data sets of the same vehicle in different environments and weather conditions, as well as that of the low accuracy of the SSD algorithm when used in detecting compartments in some scenes;

(2) The visual localization model was established, and the coordinates of the sampling point were located by color screening, image segmentation, and connecting body feature screening. The coordinates of the sampling point were converted from the image coordinate system (ICS) to the sampling plane coordinate system (SPCS), and then the robot was guided to the corresponding random point for accurate localization and sampling;

(3) A sampling verification experiment was carried out on the delivery truck on the spot. The experimental results showed that the visual localization errors of the robot in the x-direction and y-direction were less than the error margins, which means it could meet the accuracy requirements of on-site localization and sampling, and facilitate automatic sampling by a robot.

The flow chart of the study system is as follows:
2. The Proposed System

In this section, the proposed system is explained in detail. The proposed system is based on four sections: image distortion correction; SSD detection of the area of the carriage; the visual localization model; coordinate transformation, and guided robot localization and sampling. The details are shown below.

2.1. Distortion Correction

2.1.1. Lens Distortion

The ore powder transport vehicles kept at the detection site are of different sizes, and the carriages of the largest car are six to eight meters long. As shown in the Figure 2 below, the height of the pound room containing the quality inspection area is eight meters. Because the coverage of the scene to be shot is relatively large, the whole area cannot be shot with an ordinary lens, so a wide-angle lens must be adopted.

However, due to the optical system design, processing and manufacturing processes, external forces and other factors, a wide-angle lens will cause nonlinear distortion in the image captured; hence, the accurate transformation of the coordinate system cannot be achieved by using the ideal pinhole imaging model. Therefore, before the sampling area is detected in the image, the distortion parameters need to be calculated via camera calibration so as to correct the distortion of the image of the site.

Thin lens distortion refers to tilt distortion in the tangential and radial directions caused by the design of the lens itself, deviations in lens manufacturing, and errors in the processing and assembling. In general, thin lens distortion has little influence on the system.
and can be ignored; only radial distortion and centrifugal distortion are considered. Image distortion can be expressed as follows [45,46]:

\[
\begin{bmatrix}
  cR^2 & cR^4 & cR^6 & 2cr & R^2 + 2c^2 \\
  rR^2 & rR^4 & rR^6 & R^2 + 2r^2 & 2cr
\end{bmatrix}
\begin{bmatrix}
k_1 \\
k_2 \\
k_3 \\
p_1 \\
p_2
\end{bmatrix}
= \begin{bmatrix}
c_{\text{corrected}} - c \\
r_{\text{corrected}} - r
\end{bmatrix}
\tag{1}
\]

Here, \((c, r)\) is the original position of the distortion point in the ICS, and \((c_{\text{corrected}}, r_{\text{corrected}})\) is the position of the distortion point after distortion correction, considering radial distortion and centrifugal distortion.

We convert Equation (1) into \(DK = d\), \(K = [k_1 \ k_2 \ k_3 \ p_1 \ p_2]^T\), \(D = \begin{bmatrix}
cR^2 & cR^4 & cR^6 & 2cr & R^2 + 2c^2 \\
rR^2 & rR^4 & rR^6 & R^2 + 2r^2 & 2cr
\end{bmatrix}\), and \(d = [c_{\text{corrected}} - c \ r_{\text{corrected}} - r]\). Then, the distortion parameters \(K\) can be obtained by the least-squares method.

\[
K = \left(D^TD\right)^{-1}D^Td
\tag{2}
\]

2.1.2. Camera Calibration

Because the actual internal parameters of the camera are often inconsistent with expectations, in the actual use of cameras, it is necessary to determine the accurate internal parameters [47]. As shown in Figure 3 below, a 10 × 7 checkerboard was selected as the calibration object for the industrial camera. The black and white grids in the checkerboard are evenly distributed, and have good stability under different light intensities, which makes the image useful for completing the operation of corner calibration.

![Figure 3. Black and white checkerboard calibration plate.](image)

The MATLAB camera calibration toolbox is used to obtain images with different angles and different positions by changing the position of the camera or the calibration plate. Then, the corners of the calibration plate are extracted from the image, and five internal matrix parameters and five distortion parameters are solved. The parameters are optimized through the Levenberg–Marquardt algorithm. The specific process of camera calibration is as follows [48]:

1. Attach the printed calibration plate to a flat surface to ensure that it is flat;
2. Images of calibration plates at different angles and positions can be obtained by moving the camera or calibration plate. The number of calibration images is 17 (generally 15–25 [49]). During the shooting process, ensure that the calibration plate is captured completely, and the position of the calibration plate covers the image as much as possible, so as to ensure enough information is captured to calculate the projection matrix of the camera, as shown in Figure 4 below.
2.1.3. Distortion Correction Based on MATLAB Camera Calibration Toolbox

The MATLAB camera calibration toolbox was employed to read the 17 captured images of the calibration plates, as shown in Figure 5a. The corners of the 17 images were extracted via the Extract grid corners function. In the process of corner extraction, the four corners of the calibration plate need to be manually specified. After accurate calibration of the range, the actual sizes of the black and white grids of the calibration plate need to be input manually. After the input is complete, the calibration toolbox will automatically extract the corners of the calibration plate and mark them in the image, as shown in Figure 5b.

After the corners of 17 images have been extracted, the relative position and angle of the camera and the calibration plate can be intuitively determined through the Reproject images function. As shown in Figure 6a, $O_c$ on the left is the origin of the camera coordinate system, and the camera coordinate system $(O_c, X_c, Y_c, Z_c)$ is established. The different grids represent the positions of the calibration plate relative to the camera coordinate system. In Figure 6b, the calibration plate in the world coordinate system is used as a reference, and the different pyramids represent the positions of the camera relative to the calibration plate.
Figure 6. The relative positions and angles of the camera and the calibration plate. (a) Camera coordinate system. (b) World coordinate system. 

The projection error distribution of the corners extracted from images 1–17 can be obtained through the Analyze error function, as shown in Figure 7.

Figure 7. Error distribution diagram of corners.

In the 17 calibration plate images showing different positions, the maximum value of the projection error in the x-direction is less than 1.5 pixels, and the maximum value of the projection error in the y-direction is less than 2 pixels. The error is small and meets the requirements of system accuracy.

After the camera is calibrated, the internal parameters and distortion parameters of the camera are calculated through the Calibration function. The camera internal parameters table includes scale factors $f_u$ and $f_v$ ($f_u = f_x/du$, $f_v = f_y/dv$) in the x- and y-directions, the image center coordinates $c_0$ and $r_0$, and the distortion factor $a_e$. The specific parameters' values are shown in Table 1.
Table 1. Camera internal parameters table.

| Camera Internal Parameters | x-Direction  | y-Direction  |
|---------------------------|-------------|-------------|
| Focal Length (fc)         | 1167.20744  | 1166.56816  |
| ±Δfc                     | ±7.30839    | ±7.06835    |
| Principle Point (cc)      | 1017.81138  | 773.90719   |
| ±Δcc                     | ±6.56265    | ±6.01459    |
| Skew (αc)                 | 0.0000      | Angle of Pixel Axes = 90.00000 |
| ±αc                      | ±0.0000     | ±0.00000    |

$k_1$, $k_2$, and $k_3$ are radial distortion parameters, $p_1$ and $p_2$ are centrifugal distortion parameters, and the specific parameter values are shown in Table 2.

Table 2. Distortion parameters table.

| Distortion Parameter | $k_1$      | $k_2$      | $k_3$      | $p_1$  | $p_2$  |
|----------------------|------------|------------|------------|--------|--------|
| Distortion ($k_c$ and $p_c$) | -0.11795   | 0.09822    | 0.00000    | 0.00074 | -0.00124 |
| ±Δ$k_c$ and ±Δ$p_c$ | ±0.01000   | ±0.02870   | ±0.00000   | ±0.00111 | ±0.00116 |

After the camera’s internal parameters and distortion parameters are obtained, distortion correction is carried out through the initUndistortRectifyMap function and remap function of OpenCV, and the comparison of before and after distortion correction is shown in Figure 8.

![Figure 8. Comparison of before and after distortion correction. (a) Before distortion correction. (b) After distortion correction.](image)

The image captured by the industrial camera shows obvious nonlinear distortion, including obvious pincushion distortion at the edge of the image, and the carriage also shows slight barrel distortion. This will not only affect the detection accuracy of the sampling area, but also the subsequent visual localization accuracy. After distortion correction, the nonlinear distortion of the image is significantly improved, which lays the foundation for the subsequent localization accuracy.

2.2. SSD Algorithm Is Used to Detect the Carriage Area

2.2.1. SSD Algorithm

The network structure of the SSD algorithm model is shown in Figure 9. Each convolutional layer will output feature maps of the receptive fields of different sizes. On these different feature maps, the target locations and categories are trained and predicted so as to achieve multi-scale detection. This can also overcome the problems of the uncommon aspect ratio and the low recognition accuracy, and greatly improve the generalization ability [21], which is also key to the selection of SSD in this paper.
2.2.2. Experimental Process and Analysis of Sampling Area Detection

Image Acquisition

Since the industrial camera is fixed on the top and the scene is semi-enclosed, factors such as the type of vehicle, shadow, image brightness, and others have a greater impact on the detection of the sampling area. In order to improve the adaptability of the SSD algorithm to the environment, the scene was pictured at different times. The pictures of the trucks in different weather conditions are used as the training set, and some of the pictures from the training set are shown in Figure 10 below.

Data Set Pre-Processing

The training of the target detection model requires a large number of data sets for feature extraction and learning. The pictures collected on site are limited. In order to obtain a data set with sufficient training samples, brightness and contrast adjustment, noise addition, image flipping, image blurring, and other operations are performed on each picture via data augmentation [50], as shown in Figure 11 below.
Figure 11. Data augmentation renderings.

Data Set Annotation

Before model training, the sampling range in each training set image needed to be labeled with a real frame. LabelImg was used to label each image in each training set with a rectangular frame, as shown in Figure 12 below. After labeling, an XML file was generated to store the target category of the image and the position information of the rectangular frame, etc. During the training, the vector parameters in the XML file were used as the default frame in the regression process.

Figure 12. LabelImg data set annotation.

Data Conversion and Division

The deep learning framework used in the experiment is Tensorflow-GPU-1.12, which performs the conversion from XML format to tfrecord format through script files. In the process of data conversion and storage, samples in the data set are divided into a training set and a verification set (3:1), as shown in Figure 13 below.

```python
train_list = xml_list[0: int(len(xml_list) * 0.67)]
eval_list = xml_list[int(len(xml_list) * 0.67) + 1: ]
```

Figure 13. Data set division code.

2.2.3. Training Process

Before data training, the parameters related to the model training process should be set, as shown in Table 3.
### Table 3. Parameter settings of model training.

| Parameter Name        | Numerical Value | Parameter Name       | Numerical Value |
|-----------------------|-----------------|----------------------|-----------------|
| weight_decay          | 0.0005          | decay_factor         | 0.95            |
| optimizer             | Adam            | batch_size           | 8               |
| initial_learning_rate | 0.004           | num_steps            | 20,000          |

Here, weight_decay is the weight decay coefficient; optimizer is the optimizer of the model’s gradient descent; initial_learning_rate is the initial learning rate, and decay_factor is the decay factor. When approaching the optimal solution, the convergence speed of the model can be accelerated by reducing the learning rate. Batch_size is the size of each batch of data, and num_steps is the number of training steps.

The image size of the training set is 819 × 614 pixels, and the number of training cycles is 20,000, which takes 5 h and 36 min. During the training process, the change of classification loss is as shown in Figure 14a, the change of localization loss is as shown in Figure 14b, and the change of total loss is as shown in Figure 14c. With the increase in training cycles, the loss gradually converged.

![Figure 14](image-url)

**Figure 14.** Changing diagram of each loss during model training. (a) Changing process of classification loss. (b) Changing process of location loss. (c) Changing process of total loss.

After training, a model is generated that can be called. Through the generated model, the location of the sampling area of the field image can be detected, and a prediction frame can be generated.

#### 2.2.4. Regional Detection Results and Analysis

After training, the model was used in the test set for detection, and the detection effects before and after data augmentation were compared. If the prediction frame generated by the SSD algorithm exceeds the actual sampling range, this will cause security problems.
If the generated prediction frame is too small, it will not completely cover all sampling areas. Therefore, $\text{IoU}_G$ is defined as the ratio of the area of the intersection between the prediction frame and the real frame, divided by the area of the real frame. $\text{IoU}_G$ is used to measure the detection effect of the SSD algorithm in the sampling area. If the prediction frame exceeds the sampling range, the sampling range is not detected or $\text{IoU}_G < 0.8$, at which point the detection is considered a failure; otherwise, the detection is considered to be successful. The trained model performs sampling area detection on 200 test sets, and the detection results are shown in Figure 15 below.

![Graphs of detection results in test sets of different scenarios.](image)

**Figure 15.** Graphs of detection results in test sets of different scenarios. (a) Small part of the lighting scenes. (b) Most lighting scenes. (c) Cloudy scenes. (d) Night lighting scenes.
As can be seen from Table 4, before data augmentation, due to the time limit of transport vehicles, there are fewer data sets available for the same vehicle in different light intensities and contrasts, which results in two test images not within the sampling range of image detection. The prediction frame of 25 test set images exceeds the sampling range, and the prediction frame of 23 test set images \( \text{IoU}_G < 0.8 \). After data augmentation, the detection success rate increases from 74% to 92%, indicating that using data augmentation can effectively improve the detection success rate, as well as solving the problem of the low detection accuracy of some scenes using SSD (Single-Shot MultiBox Detector).

**Table 4.** Comparison of SSD model’s detections in 200 test set images.

| Sampling Range Detection Situation | Without Data Augmentation | With Data Augmentation |
|-----------------------------------|---------------------------|------------------------|
| \( \text{IoU}_G < 0.8 \)            | 23 sheets                 | 7 sheets               |
| Prediction box out of sampling range | 25 sheets               | 6 sheets               |
| No sampling range detected         | 4 sheets                  | 3 sheets               |
| \( 0.8 \leq \text{IoU}_G \leq 1 \) and not exceeding the sampling range | 148 sheets                | 184 sheets             |
| Success ratio                      | 74%                       | 92%                    |

2.3. Visual Localization Model

Due to the employed plane of sampling and the plane parallel to the camera coordinate system, conversion between the image plane coordinate system and the world coordinate system can be simplified. A visual localization model based on the bridge sampling robot, which is easy to deploy and meets the requirements of sampling accuracy, is proposed. Taking the centroid of the laser spot in the image as the reference, and combining this with the distance of the camera and the laser localization lamp, the sampling points generated in the image are translated into the SPCS, which is used to guide the sampling robot towards the randomly generated sampling points. The schematic diagram of the visual localization model is shown in Figure 16.

**Figure 16.** Schematic diagram of the visual localization model.
In the visual localization model, the conversion from the ICS to the SPCS can be achieved through the following steps.

(1) After the laser localization lamp on the top projects the green spot onto the sampling plane, through the imaging characteristics of the laser spot, it can determine the position of the green laser spot in the image, and then calculate the distance \( \Delta c \) between the center point of the image and the centroid of the spot.

(2) Based on the installation distance between the industrial camera and the laser localization lamp (\( \Delta x \) and \( \Delta c \)), the factor of conversion \( K_t \) from the ICS to the SPCS is calculated, and then the conversion from the sampling point coordinates \((c_x, r_i)\) of the ICS to the actual sampling point coordinates \((x_i, y_i)\) of the SPCS is performed.

2.3.1. Spot Detection

The laser spot has good directivity and small divergence, and is commonly used in optical localization and optical measurement [51]. As shown in Figure 17 below, whether it is cloudy or sunny, there is a large difference between the laser spot and the background. Therefore, the difference between the laser spot and the background can be used in combination with the characteristics of the laser spot to effectively detect the laser spot.

![Figure 17. The imaging diagram of the laser spot on the mineral powder plane. (a) Cloudy day. (b) Sunny day.](image)

2.3.2. Color Space Conversion

Although the color pictures stored and displayed on a computer are usually in RGB color space, RGB color space is more sensitive to light, and the color is greatly affected by brightness. Since HSV color space is more in line with human visual characteristics, the changing of various color components can be more intuitively felt in HSV color space. Moreover, in the H channel (hue channel), the color distribution is a continuous interval. Therefore, depending on the laser spot’s color, a range of hue intervals can be set to convert the image from the RGB color space to the HSV color space so as to reduce the influence of light on the detection of green laser spots. The color space conversion process is shown in Figure 18.

Both HSV and RGB use three channels to describe color information, so there is a linear transformation between the three channels of RGB and HSV. For conversion from RGB color space to HSV color space, the values of the R, G, and B channels are normalized first, the value range of each is set to \([0, 1]\), the value range of channel H is set to \([0^\circ, 360^\circ]\),
and the value range of channel S and Channel V is set to [0, 1]. Color space conversion is carried out through the following equation [52]:

\[
H = \begin{cases}
0^\circ, & \text{if } \text{max} = \text{min} \\
60^\circ \times \frac{\text{G} - \text{B}}{\text{max} - \text{min}} + 0^\circ, & \text{if } \text{max} = \text{R} \text{ and } \text{G} < \text{B} \\
60^\circ \times \frac{\text{G} - \text{B}}{\text{max} - \text{min}} + 360^\circ, & \text{if } \text{max} = \text{R} \text{ and } \text{G} \geq \text{B} \\
60^\circ \times \frac{\text{B} - \text{R}}{\text{max} - \text{min}} + 120^\circ, & \text{if } \text{max} = \text{G} \\
60^\circ \times \frac{\text{R} - \text{G}}{\text{max} - \text{min}} + 240^\circ, & \text{if } \text{max} = \text{B}
\end{cases}
\]  

(3)

\[
S = \begin{cases}
0, & \text{if } \text{max} = 0 \\
\frac{\text{max} - \text{min}}{\text{max}} = 1 - \frac{\text{min}}{\text{max}}, & \text{otherwise}
\end{cases}
\]  

(4)

\[
V = \text{max}
\]  

(5)

where \(H\) is the hue value of the pixel in the HSV color space, \(S\) is the saturation value of the pixel in the HSV color space, \(V\) is the brightness value of the pixel in the HSV color space, \(R\) is the red component of pixels in the RGB color space, \(G\) is the green component of pixels in the RGB color space, and \(B\) is the blue component of the pixel in the RGB color space. Max represents the maximum value of \(R\), \(G\), and \(B\), and min represents the minimum value of \(R\), \(G\), and \(B\).

Through the above equations, the color conversion from the RGB color space to the HSV space can be completed, and then, pixels of corresponding colors can be screened from the original image according to the actual imaging color of the laser spot.

Figure 18. RGB color space to HSV color space conversion.

2.3.3. Color Feature Screening

Since the image is stored in the computer, the computer compresses the data into one byte, in order to facilitate processing. Therefore, in OpenCV, the range of values for the three channels in the RGB and HSV images is 0–255.

Based on the color of the green laser projected onto the mineral powder plane, we set the ranges of the \(H\) channel, \(S\) channel, and \(V\) channel to ensure that the green pixels in the image can be accurately detected in the image under different lighting conditions. Based on the HSV color space model, the specific settings are shown in Table 5.

Table 5. The setting range of the three HSV channels.

| Color Channel                  | Setting Range |
|-------------------------------|---------------|
| H Channel (Hue Channel)       | (55, 77)      |
| S Channel (Saturation Channel)| (43, 255)     |
| V Channel (Value Channel)     | (46, 255)     |
Based on the setting range of the three HSV channels, a mask (object template) is generated, and a logical AND operation is performed with the original image to obtain all pixels that meet the conditions in the HSV image, as shown in Figure 19.

Figure 19. The original image after color screening.

From the partial enlargement shown on the right, the green laser spot can be screened from the HSV image based on the color characteristics, but the other green pixels in the image may be screened as well. The positions of these interfering pixels are random, so the position of the green laser spot cannot be obtained directly from the image using color features alone. It is also necessary to screen out the position of the green laser spot based on other features of the green laser spot after color screening.

2.3.4. Image Segmentation

Although green pixels can be screened out from the image based on the HSV color space, there are also green interference pixels that affect the detection of green laser spots, so the image segmentation method is adopted to separate the green laser spots from the background.

The most common methods of image segmentation include the region growing method and threshold segmentation method. In the research, the area of the green laser spot in the image is small, and the position in the image also changes with the height of the sampling plane, which is not conducive to the setting of seed pixels. In addition, the noise or uneven grayscale distribution of the image may cause voids or over-segmentation, so it is not suitable for image segmentation using the region growing method; as such, the relatively stable threshold segmentation method is used to extract the green laser spot.

Threshold segmentation methods mainly include the artificial threshold method and the maximum variance method. Whether the artificial threshold method can effectively segment the image mainly depends on whether there is enough contrast between the object and the background. However, under different lighting conditions, there are differences in the images captured by the industrial camera. The artificial threshold method cannot effectively separate the green laser spot from the background, so the maximum interclass variance method is used instead.
Assuming that the image consists of only the target object and the background, the grayscale distribution probability density of the background is \( F_1(T) \), the grayscale distribution probability density of the target object is \( F_2(T) \), and the ratio of the pixel of the target object to the pixel of the whole image is \( \varphi \), so the grayscale probability density distribution \( F(T) \) of the whole image can be expressed through the following equation [53]:

\[
F(T) = \varphi F_2(T) + (1 - \varphi) F_1(T)
\]  
(6)

Assuming that the image consists of a dark background and a bright target object, \( F_1(T) \) and \( F_2(T) \) can be represented in Figure 20. Pixels with grayscale values less than \( T_z \) are considered as the background, and pixels with grayscale values greater than \( T_z \) are considered as the target object [54].

![Grayscale value distribution of background and target object.](image)

Figure 20. Grayscale value distribution of background and target object.

After segmentation of the background and the target object via gray threshold \( T_z \), the probability of the background being mistakenly used as the target object can be calculated as [55]:

\[
E_1(T_z) = \int_{-\infty}^{T_z} F_1(T) dT
\]  
(7)

Similarly, the probability of the target object being mistakenly used as the background is:

\[
E_2(T_z) = \int_{-\infty}^{T_z} F_2(T) dT
\]  
(8)

The total probability of mis-segmentation is the sum of \( E_1(T_z) \) and \( E_2(T_z) \):

\[
E(T_z) = (1 - \varphi) E_1(T_z) + \varphi E_2(T_z)
\]  
(9)

If there is a \( T_z \) that can make \( E(T_z) \) obtain the minimum value, then \( T_z \) is the optimal threshold. The optimal threshold \( T_z \) can be calculated by deriving \( E(T_z) \) and making \( \frac{dE(T_z)}{dT_z} = 0 \); the result is as follows:

\[
\varphi F_2(T_z) = (1 - \varphi) F_1(T_z)
\]  
(10)

Assuming that \( F_1(T) \) and \( F_2(T) \) satisfy a normal distribution [56], the mean grayscale values are \( \mu_1 \) and \( \mu_2 \), and the standard deviations are \( \sigma_1 \) and \( \sigma_2 \), \( F_1(T) \) and \( F_2(T) \) can be expressed through the following equation:

\[
F_1(T) = \frac{1}{\sqrt{2\pi}\sigma_1} e^{-\frac{(T - \mu_1)^2}{2\sigma_1^2}}
\]  
(11)

\[
F_2(T) = \frac{1}{\sqrt{2\pi}\sigma_2} e^{-\frac{(T - \mu_2)^2}{2\sigma_2^2}}
\]  
(12)

Substituting (11) and (12) into (10), and taking the logarithm of both sides of the equation and simplifying them, the following equation can be obtained:

\[
AT_z^2 + BT_z + C = 0
\]  
(13)
Here, $A = \sigma_1^2 - \sigma_2^2$, $B = 2(\mu_1 \sigma_1^2 - \mu_2 \sigma_2^2)$, and $C = 2\sigma_1^2 \sigma_2^2 \ln(\sigma_2 \phi / \sigma_2(1-\phi)) + \sigma_1^2 \mu_2^2 - \sigma_2^2 \mu_1^2$.

Based on the information in the image, the parameters $\phi$, $\mu_1$, $\sigma_1$, and $\sigma_2$ are solved. Then Equation (13) is solved, and the result of the solution is the optimal threshold value.

After segmentation of the grayscale image through the maximum interclass variance method, a binary image is obtained. The foreground of the binary image contains not only the connecting body of the green laser spot, but also the connecting body of the interfering pixel, as shown in Figure 21b:

![Binary image obtained via the maximum interclass variance method. (a) Original image. (b) Binary image.](image)

The binary image contains the connecting body of the interfering pixels, so it is necessary to screen using the connecting body characteristics of the green laser spot to obtain its centroid position.

2.3.5. Connecting Body Feature Screening

Based on the features of laser spot imaging, a total of three contiguous features, namely, the center position $(x_M, y_M)$, the contiguous area feature $S$, and the contiguous circularity feature $\theta$, were selected for screening [57].

$$
\begin{align*}
\begin{cases}
 c_{\text{min}} \leq c_{Mi} \leq c_{\text{max}} \\
 r_{\text{min}} \leq r_{Mi} \leq r_{\text{max}} \\
 \theta_{\text{min}} \leq \theta_i \leq 1 \\
 S_{\text{min}} \leq S_i \leq S_{\text{max}}
\end{cases}
\end{align*}
$$

where $L$ is the long axis of the continuum; $i$ is the number of the continuum in the binarized image; $n$ is the total number of the continuum in the binarized image; $c_{\text{min}}$ and $c_{\text{max}}$ are the limit positions of the green laser spot’s center coordinates in the c-direction; $r_{\text{min}}$ and $r_{\text{max}}$ are the limit positions of the green laser spot’s center coordinates in the direction of $r$; $c_{\text{min}}$, $c_{\text{max}}$, $r_{\text{min}}$, and $r_{\text{max}}$ can be measured in the lowest sampling plane and the highest sampling plane; $S_{\text{min}}$ and $S_{\text{max}}$ are the limit area of the green laser spot’s continuum; $\theta_{\text{min}}$ represents the minimum value of roundness of the green laser spot’s continuum, and $S_{\text{min}}$, $S_{\text{max}}$, and $\theta_{\text{min}}$ can be set according to the field environment.

Through the above-described feature screening, the green laser spot can be detected in the image, the center coordinates of the green spot are output, and the detection results are marked out (with the center coordinates shown as the circle). The detection results of the green laser spot in different scenes corresponding to the image are shown in Figure 22.
All images taken in the field have been detected, and the center position of the green laser spot irradiated on the sampling plane has been accurately located in the image, providing accurate information for the subsequent conversion of the ICS to the sampling plane.

2.4. Coordinate Transform

Since the sampling depth in the z-direction of the sampling robot is controlled by the motor, and the green laser spot is projected vertically on the sampling plane, the basic coordinate system of the three-dimensional sampling robot can be simplified to a two-dimensional SPCS. The origin of the SPCS coincides with the projection point of the laser localization lamp in the sampling plane. Therefore, the coordinates of the sampling point in the SPCS can be sent to the sampling robot to guide it to the corresponding location for sampling.

The lateral distance between the center of the camera and the center of the laser localization lamp can be obtained by measurement, and in the ICS, the lateral distance between the center point of the picture and the green laser spot can also be calculated. As such, the conversion from the picture coordinate system to the SPCS can be simplified to a side conversion of two similar triangles, as shown in Figure 23.

The coordinate transformation of the sampling points is calculated as follows:

\[ K_i = \frac{\Delta c}{\Delta x} \]
\[ \begin{pmatrix} x_i \\ y_i \end{pmatrix} = K_i \begin{pmatrix} c_i \\ r_i \end{pmatrix} \quad i = 1, 2, \ldots, 16 \]  

where \( \Delta c \) is a constant in the SPCS, in pixels; \( \Delta x \) varies with the height of the sampling plane, and the higher the sampling plane, the larger \( \Delta x \), measured in mm; \( K_i \) is the conversion coefficient from the ICS to the SPCS, in pixel/mm, which is related to the height of the sampling plane; \( i \) is the serial number of the sampling point; \( (x_i, y_i)^T \) shows the sampling points in the SPCS, and \( (c_i, r_i)^T \) shows the randomly generated points in the ICS, according to the rules.
When sampling planes of different heights, Equations (15) and (16) can realize the conversion from the ICS to SPCS. The coordinates of the random sampling points can be converted and transmitted to the robot through the system, and then, the sampling robot is guided to the corresponding coordinate position for sampling.

![Coordinate conversion schematic.](image)

Figure 23. Coordinate conversion schematic.

3. Field Experiment Verification

The effectivity of the image acquisition and localization system, and the accuracy of visual localization, are verified through field experiments. The specific flow of the field experiment is shown in Figure 24.

During the experiment, the coordinates of the center of mass of the green laser spot in Figure 24d are (164,272), and the conversion coefficient $K_t = 11.9282$. According to the visual localization model, the origin of the ICS needs to be changed to the center of mass coordinates of the green laser spot; then, the coordinates are converted by $K_t$, and finally, the obtained coordinates of the sampling point are sent to the system. In order to avoid accidental errors in the data, several sampling point localization experiments were conducted on the site using different trucks, and a total of 32 sampling points were selected for comparative analysis. Some of the generated sampling points in the ICS (the coordinate origin is the center of mass of the green laser spot) and the results after coordinate transformation (the coordinate origin is the projection point of the green laser light in the sampling plane) are shown in Table 6 below.

![Distortion correction.](image)

![Detection of sampling area.](image)

Table 6. ICS (in pixels) conversion to SPCS (in mm).

| Number | ICS $(c_i, r_i)$ | Sampling Plane Coordinate System $(x_i, y_i)$ | Number | ICS $(c_i, r_i)$ | Sampling Plane Coordinate System $(x_i, y_i)$ |
|--------|----------------|---------------------------------------------|--------|----------------|---------------------------------------------|
| 1      | (458, 65)     | (5463, 775)                                | 9      | (220, 81)     | (2624, 966)                                |
| 2      | (428, 33)     | (5105, 394)                                | 10     | (220, 1)      | (2624, 12)                                 |
| 3      | (428, −62)    | (5105, −739)                               | 11     | (190, −78)    | (2266, −929)                               |
| 4      | (399, 65)     | (4759, 775)                                | 12     | (160, 33)     | (1909, 394)                                |
| 5      | (339, 49)     | (4044, 584)                                | 13     | (131, −46)    | (1563, −548)                               |
| 6      | (310, 65)     | (3698, 775)                                | 14     | (101, 65)     | (1205, 775)                                |
| 7      | (280, −62)    | (3340, −739)                               | 15     | (71, 17)      | (847, 203)                                 |
| 8      | (250, 33)     | (2982, 394)                                | 16     | (12, 17)      | (143, 203)                                 |
The difference between the actual sampling point \((x'_i, y'_i)\) and the localized sampling point \((x_i, y_i)\) is due to the image distortion caused by lens distortion, the accumulated error during the operation of the sampling robot, and the measurement error due to the flatness of the sampling plane. If the error is too large, the sampling robot may go beyond the sampling area, resulting in the failure of localization sampling, so it is necessary to verify the accuracy of the image acquisition and localization system through field experiments. After completing the sampling, the brazing rod will leave a concave column in the plane of mineral powder, as shown in Figure 25. By measuring the distance (in pixels) in the image between the center point of the concave column and the corresponding sampling point, combined with the conversion coefficient, the absolute error of each sampling point can be calculated, so the absolute error of the 16 points in Table 6 is shown in Table 7.

**Figure 24.** Experimental flow chart.
The difference between the actual sampling point $[x_i, y_i]$ and the localized sampling point $[x'_i, y'_i]$ is compared, and the analysis based on the sample point data from the experiment is shown in Figure 26 below.

| Number | Sampling Plane Coordinate System $(x_i, y_i)$ | Real Coordinate System $(x'_i, y'_i)$ | Absolute Error | Number | Sampling Plane Coordinate System $(x_i, y_i)$ | Real Coordinate System $(x'_i, y'_i)$ | Absolute Error |
|--------|---------------------------------------------|-------------------------------------|---------------|--------|---------------------------------------------|-------------------------------------|---------------|
| 1      | (5463, 775)                                | (5499, 787)                         | (36, 12)      | 9      | (2624, 966)                                | (2624, 966)                         | (0, 0)        |
| 2      | (5105, 394)                                | (5117, 394)                         | (12, 0)       | 10     | (2624, 12)                                 | (2624, 0)                           | (0, 12)       |
| 3      | (5105, −739)                               | (5105, −751)                        | (0, 12)       | 11     | (2266, −929)                               | (2234, −917)                       | (12, 12)      |
| 4      | (4759, 775)                                | (4759, 775)                         | (0, 0)        | 12     | (1909, 394)                                | (1909, 394)                        | (0, 0)        |
| 5      | (4044, 584)                                | (4068, 596)                         | (24, 12)      | 13     | (1563, −548)                               | (1575, −548)                       | (12, 0)       |
| 6      | (3698, 775)                                | (3698, 775)                         | (0, 0)        | 14     | (1205, 775)                                | (1205, 775)                        | (0, 0)        |
| 7      | (3340, −739)                               | (3340, −739)                        | (0, 0)        | 15     | (847, 203)                                 | (847, 215)                         | (0, 12)       |
| 8      | (2982, 394)                                | (2994, 394)                         | (12, 0)       | 16     | (143, 203)                                 | (143, 215)                         | (0, 12)       |

Since image distortion correction is the means to ensure the accuracy of visual localization and has a great impact on the accuracy of visual localization, the error before and after distortion correction is compared, and the analysis based on the sample point data from the experiment is shown in Figure 26 below.

![Figure 25. The concave column left after the drill rod is sampled.](image)

**Table 7. Visual localization results and actual location of the sampled brazing rod (Unit: mm).**

![Experimental comparison before and after distortion correction](image)

**Figure 26. Comparison of before and after distortion correction.**

As can be seen from the above figure, the maximum error and the average error of coordinate transformation in the x-direction and the y-direction are significantly reduced after distortion correction. The maximum error is reduced the most, with the maximum error in the x-direction reduced by 36 mm and the maximum error in the y-direction reduced by 21 mm. The experimental results verify the necessity and correctness of distortion correction in the image acquisition stage, and also contribute to the improvement of detection accuracy.

In order to facilitate the comparison of the errors between the points in the SPCS and the coordinates of the actual sampling points, the absolute error diagrams of 32 sampling points were drawn, according to Table 7 and as shown in Figure 27 below.
Due to the inaccurate localization of the sampling robot in the sampling process, resulting in sampling points being identified beyond the sampling area, according to the sampling point error range requirements, the randomly generated sampling points have a 152.78 mm error margin in the x-direction and an 83.33 mm error margin in the y-direction from the boundary of the sampling area, as shown in Table 8.

Table 8. X-direction and y-direction localization error range requirements.

| Parameters                  | X-Direction | Y-Direction |
|-----------------------------|-------------|-------------|
| Margin of error             | 152.78 mm   | 83.33 mm    |
| Error between actual and localization | ≤50 mm     | ≤50 mm      |

Here, the margin of error represents the maximum error of all randomly sampled points in the x- and y-directions that do not exceed the sampling area after the coordinate transformation. As can be seen from Figure 27, the maximum error in the x-direction is 36 mm, and the maximum error in the y-direction is 24 mm. The maximum errors in the x-direction and y-direction are much smaller than the error margin, so there will be no situation where the sampling location point is out of the sampling range. This also ensures that the error between the actual coordinates of the sampling point and the localization coordinates is less than or equal to 50 mm, which proves that the studied visual localization system can meet the requirements of industrialized practical use, and is feasible.

4. Conclusions and Further Work

4.1. Conclusions

The main focus of this paper is the perceptual localization of a sampling area. The spatial location information of the sampling point is obtained from the collected images by fusing the SSD algorithm with image processing techniques, and the sampling robot is provided with accurate localization coordinates of the sampling point. The main research work and results in this paper are summarized as follows:

1. The lens’ internal parameters and distortion parameters were obtained using the MATLAB toolbox and a calibration plate, and OpenCV is used to correct the distortion of the picture, eliminating the influence of image distortion caused by the industrial wide-angle lens. This ensures the accuracy of localization and detection;

2. By means of data augmentation, a large number of training sets within different scenes were obtained, and experiments were conducted on the test set. The accuracy of the experimental results reached over 92%, which solves the problem whereby some scenes of the SSD algorithm were not detected accurately, and ensures the accuracy of the SSD detection area;

3. A vision localization model for bridge sampling robots has been constructed, which can precisely locate the sampling point, and convert the coordinates of the localization...
point from the ICS to the base coordinate system of the sampling robot, before guiding the sampling robot to the corresponding sampling point for accurate localization and sampling;

(4) The maximum error of the localization sampling system in the x-direction is 36 mm and the maximum error in the y-direction is 24 mm. This visual localization error is much smaller than the error margin in the corresponding direction, and the localization accuracy meets the error range requirement of industrialized mineral powder sampling. Compared with the traditional conventional sampling technology, it has the advantages of shorter sampling and positioning time (originally 3–5 min for a single point, but now 8 min for a total of 16 points), stronger intelligent working ability and larger sampling range (originally the sampler can only detect half of the carriage range, but now the full range can be detected). As such, the automation of robot mineral powder localization and sampling is successfully realized.

4.2. Further Work

The data set used in the SSD model was collected on-site within a period of three months, and lacks data from other time periods. Although data augmentation has been performed to improve accuracy, the model may still fail to locate the sampling area effectively in some future weather scenarios. Therefore, for future use of the model, we must collect a large number of data sets for retraining, or choose a more effective target detection model to improve the robustness of the algorithm, as the research in the field of target detection evolves.

Author Contributions: Conceptualization, S.Y., Z.Y. and L.Z.; methodology, S.Y. and Z.Y.; validation, S.Y., Z.Y. and L.Z.; formal analysis, S.Y. and Z.Y.; investigation, S.Y., H.X. and S.Z.; writing—original draft preparation, S.Y.; writing—review and editing, S.Y., Z.Y. and L.Z.; funding acquisition, Z.Y. All authors have read and agreed to the published version of the manuscript.

Funding: This work was funded by the project ‘Intelligent Sampling Technology of China’ (XQ2018128).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Not applicable.

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

References

1. Barnewold, L.; Lottermoser, B.G. Identification of digital technologies and digitalisation trends in the mining industry. Int. J. Min. Sci. Technol. 2020, 30, 747–757. [CrossRef]
2. Tang, Y.; Zhu, M.; Chen, Z.; Wu, C.; Chen, B.; Li, C. Seismic performance evaluation of recycled aggregate concrete-filled steel tubular columns with field strain detected via a novel mark-free vision method. Structures 2021, 37, 426–441. [CrossRef]
3. Wu, F.; Duan, J.; Chen, S.; Ye, Y.; Ai, P.; Yang, Z. Multi-Target Recognition of Bananas and Automatic Positioning for the Inflorescence Axis Cutting Point. Front. Plant Sci. 2021, 12, 705021. [CrossRef] [PubMed]
4. Xue, S.; Li, X.S.; Xie, J. A New Coal Sampling System for Measurement of Gas Content in Soft Coal Seams. Appl. Mech. Mater. 2011, 121–126, 2459–2464. [CrossRef]
5. Conti, R.S.; Zlochower, I.A.; Sapko, M.J. Rapid Sampling of Products During Coal Mine Explosions. Combust. Sci. Technol. 2007, 75, 195–209. [CrossRef]
6. Yang, N.; Xie, C.; Chen, Y.; Chen, M.; Zheng, J.; Zhang, M.; Li, L. The Design of Sampling Machine for Mineral Resources. In Proceedings of the 27th International Ocean and Polar Engineering Conference, San Francisco, CA, USA, 25 June 2017; pp. 17–21.
7. Zhu, Q. Coal Sampling and Analysis Standards; IEA Clean Coal Centre: London, UK, 2014; pp. 18–37.
8. Kissell, F.N.; Volkwein, J.C.; Kohler, J. Historical perspective of personal dust sampling in coal mines. In Proceedings of the Mine Ventilation Conference, Adelaide, Australia, 2002; pp. 1–5.
9. Lecun, Y.; Bengio, Y.; Hinton, G. Deep learning. Nature 2015, 521, 436–444. [CrossRef]
10. Han, J.; Zhang, D.; Cheng, G.; Liu, N.; Xu, D. Advanced Deep-Learning Techniques for Salient and Category-Specific Object Detection: A Survey. IEEE Signal Process. Mag. 2018, 35, 84–100. [CrossRef]
11. Xu, X.; Li, Y.; Wu, G.; Luo, J. Multi-modal Deep Feature Learning for RGB-D Object Detection. Pattern Recognit. 2017, 72, 300–313. [CrossRef]
12. Ranjan, R.; Sankaranarayanan, S.; Bansal, A.; Bodla, N.; Chen, J.C.; Patel, V.M.; Castillo, C.D.; Chellappa, R. Deep Learning for Understanding Faces: Machines May Be Just as Good, or Better, than Humans. *IEEE Signal Process. Mag.* 2018, 35, 66–83. [CrossRef]

13. Chin, T.W.; Yu, C.L.; Halpern, M.; Genc, H.; Tsao, S.L.; Reddi, V.J. Domain-Specific Approximation for Object Detection. *IEEE Micro* 2018, 38, 31–40. [CrossRef]

14. Girshick, R.; Donahue, J.; Darrell, T.; Malik, J. Rich Feature Hierarchies for Accurate Object Detection and Semantic Segmentation. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Columbus, OH, USA, 23–28 June 2014; pp. 580–587.

15. Girshick, R. Fast R-CNN. In *Proceedings of the IEEE International Conference on Computer Vision (ICCV)*, Santiago, Chile, 7–13 December 2015; pp. 1440–1448.

16. Ren, S.; He, K.; Girshick, R.; Sun, J. Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks. *IEEE Trans. Pattern Anal. Mach. Intell.* 2017, 39, 1137–1149. [CrossRef]

17. Redmon, J.; Divvala, S.; Girshick, R.; Farhadi, A. You Only Look Once: Unified, Real-Time Object Detection. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Las Vegas, NV, USA, 27–30 June 2016; pp. 779–788.

18. Redmon, J.; Farhadi, A. YOLO9000: Better, Faster, Stronger. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Honolulu, HI, USA, 21–26 July 2017; pp. 7263–7271.

19. Redmon, J.; Farhadi, A. YOLOv3: An Incremental Improvement. Computer Vision and Pattern Recognition. *arXiv 2018*, arXiv:1804.02767.

20. Bochkovskiy, A.; Wang, C.Y.; Liao, H. YOLOv4: Optimal Speed and Accuracy of Object Detection. Computer Vision and Pattern Recognition. *arXiv 2020*, arXiv:2004.10934.

21. Liu, W.; Anguelov, D.; Erhan, D.;Szegedy, C.; Reed, S.; Fu, C.; Berg, A.C. SSD: Single Shot MultiBox Detector. In *Proceedings of the European Conference on Computer Vision*, Amsterdam, The Netherlands, 17 September 2016; Volume 9905, pp. 21–36.

22. Kim, J.A.; Sung, J.Y.; Park, S.H. *Comparison of Faster-RCNN, YOLO, and SSD for Real-Time Vehicle Type Recognition*; IEEE: Seoul, Korea, 2020; pp. 1–4.

23. Morera, A.; Sánchez, A.; Moreno, A.B.; Sappa, Á.D.; Vélez, J.F. SSD vs. YOLO for Detection of Outdoor Urban Advertising Panels under Multiple Variabilities. *Sensors* 2020, 20, 4587. [CrossRef]

24. Zhu, X.; Lyu, S.; Wang, X.; Zhao, Q. TPH-YOLOv5: Improved YOLOv5 Based on Transformer Prediction Head for Object Detection on Drone-captured Scenarios. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)* Workshops, Montreal, BC, Canada, 11–17 October 2021; pp. 2778–2788.

25. Jia, W.; Xu, S.; Liang, Z.; Zhao, Y.; Min, H.; Li, S.; Yu, Y. Real-time automatic helmet detection of motorcyclists in urban traffic using improved YOLOv5 detector. *IET Image Process.* 2021, 15, 3623–3637. [CrossRef]

26. Tan, S.; Lu, G.; Jiang, Z.; Huang, L. Improved YOLOv5 Network Model and Application in Safety Helmet DeTaction. In *Proceedings of the 2021 IEEE International Conference on Intelligence and Safety for Robotics*, Tokoname, Japan, 10 May 2021; pp. 1–4.

27. Deepa, R.; Tamilselvan, E.; Abrar, E.S.; Sampath, S. Comparison of Yolo, SSD, Faster R-CNN for Real Time Tennis Ball Tracking for Action Decision Networks. In *Proceedings of the 2019 International Conference on Advances in Computing and Communication Engineering (ICACCE)*, Sathyamangalam, India, 30 April 2020; pp. 1–4.

28. Zhai, S.; Shang, D.; Wang, S.; Dong, S. DF-SSD: An Improved SSD Object Detection Algorithm Based on DenseNet and Feature Fusion. *IEEE Access 2020*, 8, 24344–24357. [CrossRef]

29. Wang, X.; Hua, X.; Xiao, F.; Li, Y.; Hu, X.; Sun, P. Multi-Object Detection in Traffic Scenes Based on Improved SSD. *Electronics 2018*, 7, 302. [CrossRef]

30. Li, J.; Hou, Q.; Xing, J.; Ju, J. SSD Object Detection Model Based on Multi-Frequency Feature Theory. *IEEE Access 2020*, 8, 82294–82305. [CrossRef]

31. Li, Y.; Dong, H.; Li, H.; Zhang, X.; Zhang, B.; Xiao, Z. Multi-block SSD based on small object detection for UAV railway scene surveillance. *Chin. J. Aeronaut.* 2020, 33, 1747–1755. [CrossRef]

32. Fu, C.; Liu, W.; Ranga, A.; Tyagi, A.; Berg, A.C. Dssd:Deconvolutional single shot detector. Computer Vision and Pattern Recognition. *arXiv 2017*, arXiv:1701.06659.

33. Dai, J.; Li, Y.; He, K.; Sun, J. R-fcn: Object detection via region-based fully convolutional networks. In *Advances in Neural Information Processing Systems 29, Proceedings of the Annual Conference on Neural Information Processing Systems 2016*, Barcelona, Spain, 5–10 December 2016; The MIT Press: Cambridge, MA, USA, 2016; pp. 379–387.

34. Lawrence, S.; Giles, C.L.; Tsoi, A.C. What size neural network gives optimal generalization? convergence properties of backpropagation. In *Computer Science*; University of Maryland: College Park, MD, USA, 1998; pp. 1–35.

35. Leng, J.; Liu, Y. An enhanced SSD with feature fusion and visual reasoning for object Detection. *Neural Comput. Appl.* 2019, 31, 6549–6558. [CrossRef]

36. Jeong, J.; Park, H.; Kwak, N. Enhancement of SSD by concatenating feature maps for object detection. In *Computer Science*; University of Maryland: College Park, MD, USA, 2017; pp. 1–35.

37. Li, Z.; Zhou, F. FSSD: Feature fusion single shot multibox detector. Computer Vision and Pattern Recognition. *arXiv 2017*, arXiv:1712.00960.
38. Al-Jubouri, Q.; Al-Azawi, R.J.; Al-Taee, M.; Young, I. Efficient individual identification of zebrafish using Hue/Saturation/Value color model. *Egypt. J. Aquat. Res.* 2018, 44, 271–277. [CrossRef]

39. Harasthy, T.; Ovseník, L.; Turán, J. Detector of Traffic Signs with using Hue-Saturation-Value color model. *Carpathian J. Electron. Comput. Eng.* 2013, 2, 21–25.

40. Wandi, D.; Fauziah, F.; Hayati, N. Deteksi Kelayuan Pada Bunga Mawar dengan Metode Transformasi Ruang Warna Hue Saturation Intensity (HSI) dan Hue Saturation Value (HSV). *J. Media Inform. Budidarma* 2021, 5, 308–316. [CrossRef]

41. Cantrell, K.; Erenas, M.M.; de Orbe-Paya, I.; Capitán-Vallvey, L.F. Use of the Hue Parameter of the Hue, Saturation Value Color Space As A Quantitative Analytical Parameter for Bitonal Optical Sensors. *Anal. Chem.* 2010, 82, 531–542. [CrossRef]

42. Wang, Z.; Jensen, J.R.; Im, J. An automatic region-based image segmentation algorithm with applications for remote sensing applications. *Environ. Model. Softw.* 2010, 25, 1149–1165. [CrossRef]

43. Ojeda, S.; Vallejos, R.; Bustos, O. A new image segmentation algorithm with applications to image inpainting. *Comput. Stat. Data Anal.* 2010, 54, 2082–2093. [CrossRef]

44. Chouhan, S.S.; Kaul, A.; Singh, U.P. Image Segmentation Using Computational Intelligence Techniques: Review. *Arch. Comput. Methods Eng.* 2018, 26, 533–596. [CrossRef]

45. Wang, J.; Shi, F.; Zhang, J.; Liu, Y. A new calibration model of camera lens distortion. *Pattern Recognit.* 2008, 41, 607–615. [CrossRef]

46. Tang, Y.; Li, L.; Feng, W.; Liu, F.; Zou, X.; Chen, M. Binocular vision measurement and its application in full-field convex deformation of concrete-filled steel tubular columns. *Measurement* 2018, 130, 372–383. [CrossRef]

47. Chen, M.; Tang, Y.; Zou, X.; Huang, K.; Li, L.; He, Y. High-accuracy multi-camera reconstruction enhanced by adaptive point cloud correction algorithm. *Opt. Lasers Eng.* 2019, 122, 170–183. [CrossRef]

48. Remondino, F.; Fraser, C. Digital camera calibration methods Considerations and comparisons. *ISPRS Comm. V Symp. Image Eng. Vis. Metrol.* 2006, 36, 266–272.

49. Zhang, Z. A Flexible New Technique for Camera Calibration. *IEEE Trans. Pattern Anal. Mach. Intell.* 2000, 22, 1330–1334. [CrossRef]

50. Buslaev, A.; Iglovikov, V.I.; Khvedchenya, E.; Parinov, A.; Druzhinin, M.; Kalinin, A.A. Albumentations: Fast and flexible image augmentations. *Information* 2020, 11, 125. [CrossRef]

51. Meško, M.; Toth, Š. Laser Spot Detection. *J. Inf. Control. Manag. Syst.* 2013, 11, 35–42.

52. Jayashree, R.A. RGB to HSI color space conversion via MACT algorithm. In Proceedings of the 2013 International Conference on Communication and Signal Processing, Melmaruvathur, India, 3–5 April 2013; pp. 561–565.

53. Qi, Q.; Tian, Y.; Han, L. An improved image segmentation algorithm based on the maximum class variance method. In Proceedings of the MATEC Web of Conferences 2020, Beijing, China, 27–29 November 2020; Volume 309, pp. 1–4.

54. Felzenszwalb, P.F.; Huttenlocher, D.P. Efficient Graph-Based Image Segmentation. *Int. J. Comput. Vis.* 2004, 59, 167–181. [CrossRef]

55. Liu, J.F.; Cao, X.L.; Xu, J.; Yao, Q.L.; Ni, H.Y. A new method for threshold determination of gray image. *Geotech. Geophys. Geo-Energy Geo-Resour.* 2020, 6, 72. [CrossRef]

56. Liu, L.; Yang, N.; Lan, J.; Li, J. Image segmentation based on gray stretch and threshold algorithm. *Optik* 2015, 126, 626–629. [CrossRef]

57. Guo, F.; Li, W.; Tang, J.; Zou, B.; Fan, Z. Automated glaucoma screening method based on image segmentation and feature extraction. *Med. Biol. Eng. Comput.* 2020, 5, 2567–2586. [CrossRef] [PubMed]