Photoelectric Target Detection Algorithm Based on NVIDIA Jeston Nano

Shicheng Zhang¹, Laixian Zhang²*, Huayan Sun² and Huichao Guo²

¹ Graduate School, Space Engineering University, Beijing 101416, China
² Department of Electronic and Optical Engineering, Space Engineering University, Beijing 101416, China
* Correspondence: zhanglaixian@126.com; Tel.: +86-1366-133-8085

Abstract: This paper proposes a photoelectric target detection algorithm for NVIDIA Jeston Nano embedded devices, exploiting the characteristics of active and passive differential images of lasers after denoising. An adaptive threshold segmentation method was developed based on the statistical characteristics of photoelectric target echo light intensity, which effectively improves detection of the target area. The proposed method’s effectiveness is compared and analyzed against a typical lightweight network that was knowledge-distilled by ResNet18 on target region detection tasks. Furthermore, TensorRT technology was applied to accelerate inference and deploy on hardware platforms the lightweight network Shuffv2_x0_5. The experimental results demonstrate that the developed method’s accuracy rate reaches 97.15%, the false alarm rate is 4.87%, and the detection rate can reach 29 frames per second for an image resolution of 640 × 480 pixels.

Keywords: photoelectric targets; lightweight network; threshold segmentation; knowledge distillation; TensorRT acceleration

1. Introduction

Pinhole cameras, miniature cameras, and other photoelectric equipment pose a severe information leakage risk that may impact society and business in bringing substantial property losses and severe threats to personal safety. This can be addressed by using the cat-eye effect of cameras [1] for detection by using an incident-light irradiation camera that traces back the original camera light-path and the reflected light intensity of the diffuse reflected light, which has 2–4 orders of magnitude higher intensity. The spot generated on the photoelectric sensor or slicing board of the focal plane inside the camera due to the reflection of the cat-eye effect is called the cat-eye target. This method is independent of the camera’s operation and the transmission of wireless signals and only depends on the camera’s physical characteristics. Thus, using laser-active imaging [2] is one of the most effective ways to detect conventional cameras.

In the early years of the discipline, detection relied on the photoelectric target’s optical characteristics, primarily including the target’s light cross-sectional characteristics [3], the optical components’ characteristics [4], and the polarization characteristics [5]. Nevertheless, their detection rate is still low. The great success of AlexNet [6], VGGNet [7], and ResNet [8] since 2010 has set off an upsurge in convolutional neural network research. The convolutional neural network [9] has a multi-level feature extraction ability, widely used in computer vision and natural language processing. Some scholars suggest using deep learning to extract feature information [10] of optoelectronic targets. For instance, Ke [11] designed a fully automatic camera detection and recognition system, which combines machine learning and neural network methods to identify surveillance camera equipment effectively. Liu [12] introduced a photoelectric target recognition algorithm and a detection system based on convolutional neural networks to detect indoor micro-cameras using classified networks. Huang [13] used the improved YOLOv3 model to identify micro-cameras in a single frame. Nevertheless, the above methods are limited by the high computing
power of the hardware affecting detection speed and performance, resulting in the detection equipment having an excessive volume and thus poor portability limiting its applicability.

Spurred by the deficiencies mentioned above, this paper develops a photoelectric target detection algorithm for the NVIDIA Jeston Nano, a small, embedded device with certain computing power limitations. In order to ensure the good real-time performance of the detection process, an adaptive threshold segmentation method based on the statistical characteristics of the photoelectric target echo light intensity is proposed during image preprocessing [14]. This strategy enhances the algorithm’s performance on the embedded device by reducing complexity. Considering deep learning [15], the prediction accuracy of lightweight networks is improved by knowledge distillation [16], and the model inference speed is greatly improved using TensorRT technology [17]. Hence, the entire algorithm process is suitable for embedded devices with small computing power [18], affording real-time detection of photoelectric targets in real-world scenarios.

The remainder of this article is organized as follows. Section 2 describes the equipment and processes, datasets, and experimental environments for data collection. Section 3 proposes an adaptive threshold segmentation method based on the statistical characteristics of photoelectric target echo light intensity. Additionally, it introduces image preprocessing, target region discrimination, and model acceleration. Section 4 analyzes the experimental results and verifies the effectiveness and real-time nature of the algorithm’s detection of photoelectric targets. Finally, Section 5 concludes this work.

2. Materials and Equipment

2.1. Data Acquisition Equipment and Processes

Imaging methods can be divided into active and passive imaging, depending on whether or not they use active light sources to irradiate the target and its background. Active imaging exploits the illumination light source (typically a laser) to irradiate the target and its background, with the advantage of enhancing the target information and not being susceptible to environmental factors. Passive imaging utilizes the natural light reflecting on the target and background, affording rich acquisition information and easy acquisition.

Experimentally, a photoelectric target laser active imaging detection system was built to acquire active and passive images (Figure 1). The developed setup included a laser with a wavelength of 532 nm and an industrial CCD camera as the detector, acquiring images of 640 × 489 pixels. The data collected by the photoelectric target involved a conventional camera with a diameter of 20 mm, and the shooting distance was 60–80 m.

![Figure 1. The data acquisition process of photoelectric target laser active imaging detection system.](image-url)

During image acquisition, the laser and detector were controlled by synchronous triggers. The detector acquired images every time the laser changed its operating state, and the detector’s operating frequency was twice that of the laser, ensuring smooth acquisition of active and passive images.
2.2. Dataset

The datasets in this experiment include exclusively authentic shots. During the experiment, highly reflective substances such as folded tin foil, plastic bottles, and metals were added to simulate false targets and increase the background complexity. In total, 2540 active and 2540 passive images were collected with the setup illustrated in Figure 1. Based on the characteristics of the active image enhancement target information, the dataset was collected and labeled in the active image (Figure 2). The camera lens’ reflected flare (the red area in the image) is the dataset’s true target. Specks of light reflected (green areas in the image) by targets, such as metal near the window and stones on the road, serve as false targets. The 2540 active images containing photoelectric targets from outdoor scenes were cropped, and 6083 dataset true target images and 6083 dataset false target images were created. After cropping, each image was of size 20 × 20 pixels, which was increased to 64 × 64 when entering the network.

![Figure 2. Sample images of a partial dataset. (a) Active image labeled with a true target in the dataset, (b) part of the true target dataset, and (c) part of the false target dataset.](image)

2.3. Experimental Environment

The hardware employed to train the deep learning model and perform knowledge distillation was Inter (R) Core (TM) i7-9700F CPU@3.00GHz, GTX TITAN XP with 12 GB GPU memory, and Windows 10 operating system with 64 GB system memory. The testing environment and TensorRT acceleration utilized an NVIDIA Jeston Nano 4 GB embedded development board, with a quad-core ARM Cortex-A57 MPCore processor as CPU and a NVIDIA Maxwell architecture as GPU, equipped with 128 NVIDIA CUDA® cores, providing 472 GFlops of computing performance.

3. Methods

The developed algorithm flowchart is illustrated in Figure 3. Specifically, the acquired active and passive images undergo differential processing and thus are converted to differential images. In the latter images, the proposed adaptive threshold segmentation method is applied based on the statistical characteristics of the echo light intensity of the photoelectric target within the differential image. Using the above method and the shape measurement criterion of the photoelectric target, the target within the suspected target area is judged by the validity, and the screened target area was mapped to the active image as a candidate region for retention. The candidate region in the active image is input to the convolutional neural network for discrimination to obtain accurate photoelectric target information. Finally, the discriminant result is marked as the true target on the active image. However, utilizing convolutional neural networks directly on embedded devices is not processing-efficient, as it does not meet the real-time requirements for hardware platform deployment. Therefore, the network model must be accelerated, which is accomplished by developing a lightweight network after applying knowledge distillation and TensorRT technology. This strategy accelerates the inference on the lightweight network and affords its
deployment on the hardware platform while meeting the real-time detection requirements.

![Algorithm Flowchart of This Article](image)

Figure 3. Algorithm flowchart of this article.

This chapter contains three parts: image preprocessing, candidate region discrimination, and model acceleration. The image preprocessing part proposes an adaptive threshold segmentation method based on the statistical characteristics of the cat-eye target echo light intensity. This segmentation scheme and the cat-eye target shape measurement criterion are used to segment the target and the background while eliminating image noise. This strategy eliminates pseudo-targets and provides the basis for subsequent target identification works. The candidate region discrimination part accurately classifies the candidate region obtained from the convolutional neural network, and the model acceleration part solves the problem of the algorithms not meeting the real-time deployment on NVIDIA Jeston Nano, by utilizing knowledge distillation and TensorRT acceleration.

3.1. Image Preprocessing

Initially, the acquired active and passive images undergo morphological processing. Since the photoelectric target appears as a circle in the image and occupies a small number of pixels, the Tophat transformation using the circle as a structural element can suppress the background factor and enhance the outline of the small target effect. Given that the active image can enhance the target’s pixel intensity, the passive image has rich information acquisition, and the image difference operation is used to eliminate the influence of background factors on target recognition.

\[
I_{sub} = I_a - \frac{[I_a + (N-1)I_p]}{N} \tag{1}
\]
where $I_a$ and $I_p$ are the active and passive images, respectively, and $N$ is the total pixel value of the grayscale image. Figure 4 depicts a set of images and their image differentiation output.

![Figure 4. Image differentiation. (a) Active image, (b) passive image, and (c) differential image.](image)

The photoelectric target irradiates the photoelectric device with an active light source, and due to the principle of reversibility of the optical path, the incident light returns along the original optical path at the focal plane within the photoelectric device. This property of the optoelectronic devices is also known as the “cat’s eye effect”. As illustrated in Figure 4a, other spurious targets are also detected when the camera is detected. In that case, the echo power of the photoelectric target and the diffuse reflection is theoretically analyzed. The echo spot area received by the detector can be expressed as

$$S = \pi \left( \frac{R\theta'}{2} \right)^2$$  \hspace{1cm} (2)

where $\theta'$ is the reflex angle of the photoelectric target. The photoelectric target echo power $P$ received by the light-sensitive surface at the camera’s focal plane has an effective area $A'$. 

$$P = p_1 \frac{4\tau^2_1 \tau^2_2 \tau_1 A_s A_r \rho}{\pi \theta^2 R^2 S}$$  \hspace{1cm} (3)

where $p_1$ is the emitted laser power, $\tau$ is atmospheric transmittance, $\tau_1$ denotes the transmittance of the emission optics system, $\tau_2$ is the photoelectric target optoelectronic device transmittance, and $\tau_1$ is the receiving optical system. $A_s$ is the photoelectric target that has a limited aperture receiving area transmittance, $A_r$ denotes the echo detector receiving area, $\theta$ is the beam dispersion angle of the emitted laser, and $R$ is detection range. The diffuse reflection for a small target echo power $P_{rs}$ is given by

$$P_{rs} = p_1 \frac{2\tau^2_1 \tau_2 A_r \rho \tau_{rs} A_{rs} \cos \theta}{\pi \theta^2 R^4}$$  \hspace{1cm} (4)

where $\rho_{rs}$ is the reflectivity of a diffused small target, $A_{rs}$ denotes the effective cross-sectional area of a diffuse reflection small target, and $\theta$ is the diffuse reflection of the laser incidence’s angle at a small target. Comparing Equation (3) with Equation (4) yields the photoelectric target echo power and the diffuse small target echo power:

$$\frac{P}{P_{rs}} = \frac{8\tau^2_1 A_s \rho}{\theta^2 A_{rs} \rho_{rs} \cos \phi}$$  \hspace{1cm} (5)

where $\theta'$ is the echo reflection angle. In practical application scenarios, the photosensitive surface reflectance $\rho$ of the photoelectric target is 0.3, the reflectivity of most diffuse targets is between 0.2–0.5, and the light wave transmittance $\tau_2$ of the photoelectric system is above 0.95. Under the condition that the target’s effective area does not significantly vary and
the light is incident, the photoelectric target echo power and the diffuse small target echo power are approximated as follows:

$$\frac{P}{P_{rs}} = \frac{8}{\theta^2}$$  \hspace{1cm} (6)

When comparing the photoelectric target echo power with the diffuse small target echo power, their ratio is constant, unaffected by the operating distance changes. Therefore, the echo power intensity of the photoelectric target is greater than the echo power intensity of the small target’s diffuse reflection. This characteristic is exploited as the theoretical basis for distinguishing true and false targets.

However, minor noise exists in the difference image, primarily from the strong noise and edge features in the active image’s background. Therefore, the difference image is further processed to solve this problem. The camera generates a strong echo signal due to the “cat’s eye” originating from the laser’s active detection, which has a mountain peak shape, where the peak pixel value is relatively the average in the image gray value transformation. In other words, the low pixel value of the background wraps the characteristics of the target’s high pixel values. Although the high-brightness pseudo-target area has a mountain peak shape in the image gray value transformation, its peak value pixels are not evenly distributed but present an irregular and uneven shape (Figure 5).

![Figure 5. Three-dimensional view of the pixel values belonging to real and pseudo targets. (a) True target local pixel value and (b) partial pseudo-target local pixel value.](image)

The real target represented in Figure 5a has a single peak identified with a smooth mountain peak straight up and down the mountain. The detector’s light saturation causes the smoothness of its summit. Opposing, the pseudo-target in Figure 5b has a continuous multi-peak feature. The area around the peak fluctuates more violently than the true target due to the changing pixel values between the pseudo-target and the surrounding pixels affected by diffuse reflection. These two characteristics are used to judge a potential target.

The experiment was mainly oriented to the NVIDIA Jetson Nano hardware platform. The traditional filtering and Fourier transforms impose a substantial computational burden, not guaranteeing an acceptable real-time performance to small computing power equipment. Therefore, the threshold segmentation method is selected, which has minimum computational requirements. Despite that, the photoelectric target is not effectively segmented, as according to the statistical analysis of the differential image data, the pixel proportion of the photoelectric target is small. As illustrated in Figure 6, most of the pixels in the differential image are distributed in the low-brightness area. Thus, by analyzing the histogram of characteristics, an adaptive threshold segmentation method is proposed based on the statistical characteristics of the echo light intensity of the optoelectronic target.
Figure 6 depicts the different images with grayscale histograms of 640 × 480 pixels in the outdoor environment. Figure 6a contains many low pixel regions, and the corresponding histogram distribution is not uniform, indicating that the difference between the background and the grayscale target values is too large. We divide the difference image into low and high luminance regions to solve this problem. The former region does not affect the threshold segmentation, as only the high luminance region is considered during threshold segmentation. Dividing the gray value between the low and the high luminance region is particularly important. Consider Figure 6 as an example: in the differential image (a), the maximum gray value of the Target A region (pseudo-target area), the Target B (true-target area) region are 198 and 199, respectively. Considering the actual situation, the high brightness reflector gray value is greater than the photoelectric target gray value. Thus, we experimentally found that splitting the high and low brightness region gray values based on the entire image’s highest 1% gray value of the first-pixel point is the most appropriate choice. This value was then used as a threshold to segment the high brightness region:

\[
T = \frac{m \times I_{\text{max}} + I_B}{m + 1}
\]

where \(I_{\text{max}}\) is the largest pixel value point in the image, \(k\) is a constant, \(m\) represents the maximum gray value in the image, and \(I_B\) is the difference degree of the background of the high-intensity region. When \(I_B\) is small, the high-intensity area has fewer false targets, and \(m\) can be approximated as the difference degree between the photoelectric target and the background of the high-intensity area. The greater the difference degree between the photoelectric target and the high-intensity background while solving for the threshold \(T\), the higher the weight of \(m\), and the more that the threshold \(T\) tends to be the maximum gray value. The comparison after threshold segmentation is presented in Figure 6c.

To further improve screening accuracy, the shape characteristics of the photoelectric targets are analyzed, and the shapes of photoelectric targets in different cameras and sites in the dataset are counted. The dataset is 60–80 m apart in an outdoor environment using a 532 nm laser and CCD industrial camera to shoot a conventional camera with a diameter of 20 mm. The image size acquired is 640 × 480 pixels, the photoelectric target is a solid area on the image after threshold segmentation, and the pixel area is between 3 × 3 and 6 × 6. The optical aperture of the photoelectric equipment is mostly circular. Therefore, the shape of the photoelectric target in the image echo signal was considered as solid and approximately circular, as illustrated in Figure 7.
According to the shape characteristics analyzed in the above experiments, the final candidate region was removed from the deformed region with a length-to-width ratio greater than 2:1 and the spot region with an area of fewer than three pixels. The shape measurement before and after the comparison chart is depicted in Figure 8, where the real photoelectric targets are marked with red boxes, parameter A represents the area of pixels occupied, and parameter B represents the aspect ratio. The spot inside the white circle is deleted because it does not meet the shape measurement criteria. Only the spot inside the red circle is retained as a candidate area for the following process.

Figure 8. Comparison of before and after shape screening (a) image after threshold segmentation and (b) image after shape measurement.

3.2. Candidate Region Discrimination

The entire preprocessing stage can simply and effectively remove a part of the false area. Considering the algorithm’s complexity, the proposed method is more friendly to
low-computing devices. Since the photoelectric target occupies fewer pixels in the image, and its optical characteristics that can be extracted are limited, it is difficult to test the photoelectric target in all scenarios accurately. Therefore, a convolutional neural network is used to identify candidate regions above to improve the algorithm’s detection ability.

In-depth studying of the convolutional neural networks reveals that increasing the network’s depth introduces more nonlinear features, making the network difficult to train. Thus, He proposed the residual module contained in ResNet [8] to solve this problem. Therefore, ResNet is selected as the feature extraction network and the teacher network for the following knowledge distillation part (see the Experimental validation in Section 4.2). The following analysis and experimental verification are made based on the ResNet layers:

1. The large difference in the number of layers between the teacher and the student models in knowledge distillation has a negative effect on the distillation results [20]. Since the hardware platform has small computing power, the student model will be a lightweight network, so the teacher model can only be a low-level network, such as ResNet18, ResNet34, and ResNet50.

2. Since the image pixel of this training set is small and the target feature information is relatively single, increasing the number of convolutional layers may discard part of the acquired low-level feature information. The prediction accuracy of ResNet18, ResNet34, and ResNet50 after training is 99.54%, 99.17%, and 98.72%, respectively.

By analyzing the experimental data and considering the relevant theory, ResNet18 was selected as the feature extraction network and teacher network of the subsequent knowledge distillation part. Through this network, the candidate region after image preprocessing can be discriminated, and if the discriminant result is the real target, the detection result can be obtained by marking the region on the original image. The candidate area may simply be discarded if the discriminant result is a false target.

3.3. Model Acceleration

Due to the hardware limitations of embedded devices, deploying the ResNet18 model directly on the device will lead to high execution times and, thus, poor real-time performance. Thus, the ResNet18 must undergo a knowledge distillation process, and TensorRT technology can be used to accelerate reasoning, effectively solving the above problems.

In 2014, Hinton [16] from Google LABS first proposed the idea of complete knowledge distillation and experimentally verified its feasibility and the effectiveness of convolutional neural network compression on MNIST data sets. In the probability distribution of the output of a well-trained network model, the probability of the error category is generally small, and its relative probability hides the characteristic information that only the real tags of 0 and 1 do not have. Therefore, the temperature coefficient $T$ is added to the output layer of Softmax to smooth the probability distribution of the network’s output. The output is called soft target, which guides the students’ network training and the real label. The corresponding loss function $J_{KD}$ is obtained as follows, where Equation (10) is a supplementary description of Equation (9):

$$J_{KD} = J_{CE}(y_{true}, p) + \gamma T^2 J_{CE}(\hat{p}_l, \hat{q}_l)$$

(9)

$$\hat{q}_l = \frac{\exp(\frac{z}{T})}{\sum \exp(\frac{z}{T})}$$

(10)

Neural networks typically generate class probabilities using a softmax output layer, which normalizes, $z$, to probability $\hat{q}_l$. $J_{CE}(y_{true}, p)$ represents the cross-entropy between the predicted output of the student network and the true label, and $J_{CE}(\hat{p}_l, \hat{q}_l)$ is the cross-entropy between the predicted output after smoothing by the student network and the teacher network. Also, $\gamma$ is a hyperparameter that adjusts the ratio of the two loss functions, and because the cross-entropy is smoothed by hyperparameter $T$, its gradient will become the original $1/T^2$ during back-propagation. To preserve the scale of its gradient consistent
with the scale of the cross-entropy corresponding to the true label, it is necessary to multiply the smoothed cross-entropy by $T^2$.

TensorRT Quantization technology [21] is a high-performance neural network reasoning optimizer launched by NVIDIA that provides low-delay and high-throughput deployment reasoning for deep-learning applications. It can accelerate reasoning for super-large-scale data centers and embedded or autonomous-driving platforms. It supports deep learning for most frameworks for fast and efficient deployment reasoning. TensorRT relies on Computeunified Device Architecture (CUDA), including an industry-level library explicitly created for deep-learning model deployment. It optimizes trained neural network models, usually 32-bit or 16-bit data. According to the programmability of CUDA, TensorRT can optimize increasingly complex and diverse deep-learning neural networks.

4. Results

4.1. Analysis of Threshold Segmentation Results

The detection performance of the proposed threshold segmentation method was experimentally validated and challenged against other threshold segmentation methods. The corresponding results are illustrated in Figure 9.

Figure 9(a1,b1) illustrate the active and passive images of a single photoelectric target with false targets, while Figure 9(a2,b2) depict the active and passive images of multiple photoelectric targets in a simple environment. The real photoelectric targets are marked with red boxes in the active image. The experimental comparison reveals that the proposed method has more advantages than traditional methods for photoelectric target segmentation. In a simple environment, the segmentation effect is ideal, while when there are false targets such as high brightness reflectors, only a part of the false targets can be screened, and the remaining must be further discriminated.

Figure 9 reveals that the developed method has an appealing effect on suppressing background factors and enhances the contrast of photoelectric targets well, providing convenient conditions for the deep learning model to discriminate the target area. This is especially important as it compensates for the algorithmic complexity of embedded platforms with less computational power.

4.2. Analysis of Knowledge Distillation

This section conducts experimental verification and analysis of the teacher and student network selection to achieve a better knowledge-distillation effect. Regarding the teacher network selection, three classical neural networks (VGG16, AlexNet, and Resnet18) were selected to train the dataset. The top-1%, parameter amount, calculation amount, and average inference time of 30 inferences in NVIDIA Jeston Nano are reported in Table 1 (- indicates that the resources occupied by NVIDIA Jeston Nano are too large to run).

Table 1. Candidate teacher network; top-1%, parameter quantity, calculation amount, and inference time in NVIDIA Jeston Nano.

| Network       | Top-1% | Parameter Quantity (M) |
|---------------|--------|------------------------|
| VGG16         | 99.94% | 134.27                 |
| AlexNet       | 99.74% | 57.00                  |
| ResNet18      | 99.54% | 23.51                  |

Table 1 shows that the prediction accuracy of all competitor teacher networks is almost the same, while only the computational burden and the parameter cardinality vary, resulting in different inference times on the NVIDIA Jeston Nano device. Since this cannot be a critical factor in selecting the teacher network, a comparative experiment of knowledge distillation between the teacher and student networks was conducted. The teacher network remained one of the above three classic neural networks, and the student network was one of the lightweight networks: Shuffv2 [22], Squeezent [23], GhostNet [24], and Con-
Neural networks typically generate class probabilities using a softmax output layer, as depicted in Equation (1). The corresponding results are illustrated in Figure 9.

$$\text{softmax}(x) = \frac{e^x}{\sum_{i=1}^{n} e^{x_i}}$$

Figure 9. Comparison between the proposed and traditional methods: (a1,a2) active image, (b1,b2) passive image, (c1,c2) maximum entropy threshold segmentation, (d1,d2) iterative threshold segmentation, (e1,e2) Otsu-based threshold segmentation, (f1,f2) proposed threshold segmentation.

denseNet2 [25]. The parameters of top-1%, parameter quantity, calculation amount, and average inference time of 30 inferences in NVIDIA Jeston Nano after knowledge distillation on the four lightweight networks are reported in Table 2.
Table 2. Lightweight network; top-1% before and after knowledge distillation, the parameters cardinality, the amount of calculation, and the inference time in NVIDIA Jeston Nano.

| Network         | Top-1%   | KD-VGG16 Top-1% | KD-AlexNet Top-1% | KD-ResNet18 Top-1% | Parameter Quantity (M) | Calculated Amount (M) | Inference Time (s) |
|-----------------|---------|-----------------|-------------------|------------------|------------------------|-----------------------|-------------------|
| Shuffv2_x0_5    | 98.84%  | 98.09%          | 97.79%            | 99.05%           | 0.34                   | 2.95                  | 0.0783            |
| Shuffv2_x1_0    | 99.26%  | 98.69%          | 98.84%            | 99.63%           | 1.26                   | 11.62                 | 0.0926            |
| Shuffv2_x1_5    | 99.50%  | 99.45%          | 99.71%            | 99.73%           | 2.48                   | 24.07                 | 0.0986            |
| Shuffv2_x2_0    | 99.67%  | 99.44%          | 99.59%            | 99.73%           | 5.35                   | 47.62                 | 0.1134            |
| Squeezent1_0    | 97.86%  | 88.34%          | 63.18%            | 97.94%           | 0.73                   | 41.74                 | 0.1134            |
| Squeezent1_1    | 91.20%  | 96.03%          | 81.81%            | 95.23%           | 0.72                   | 16.05                 | 0.0682            |
| GhostNet        | 97.94%  | 99.19%          | 99.04%            | 98.99%           | 3.90                   | 14.26                 | 0.0910            |
| CondenseNetv2   | 95.12%  | 98.54%          | 96.73%            | 98.69%           | 7.26                   | 169.0                 | -                 |

Table 2 shows that the accuracy improvement of the lightweight network after knowledge distillation is more significant than Resnet18. Thus, ResNet18 was selected as the teacher network for knowledge distillation. Since the lightweight network model was to eventually be deployed on the NVIDIA Jeston Nano embedded development device, we also analyzed the process of selecting the lightweight network. To ensure that the algorithm is executed in real-time, the inference time should be considered first in the actual deployment. Thus, Squeezent1_1, Squeezent1_0, and Shuffv2_x0_5 have advantages over other lightweight networks. Therefore, only Squeezent1_0 and Shuffv2_x0_5 were selected based on the number of parameters and the computational load on the hardware platform. Also, the prediction accuracy had a more significant impact on the results. Thus, this index is also one of the factors to be considered. Shuffv2_x0_5 had a higher prediction accuracy of nearly 5% compared to Squeezent1_0 and had a better detection performance. Therefore, the Shuffv2_x0_5 lightweight network is more appropriate for deployment on the hardware platform. Figure 10 depicts the final network structure of the teacher and student models after a comparative analysis of the experimental data.

4.3. Analysis of the Impact of Preprocessing on the Algorithm

This section verifies the effectiveness of the adaptive threshold segmentation method on the entire algorithm process. The segmentation scheme is based on the statistical charac-
teristics of the photoelectric target echo light intensity. This section compares the number of candidate regions extracted by the threshold segmentation method and analyzes the effectiveness of the threshold segmentation method on the entire algorithm process. Due to the poor effect of some traditional threshold methods (see Section 4.1), this section only compares the screening detection area considering the Otsu threshold segmentation method. The candidate areas of 100 groups of active and passive images collected from outdoor scenes were selected, including 123 real targets. The shape characteristics of the candidate areas were screened consistently during the experiment, and only the threshold segmentation method was compared. The corresponding experimental results are illustrated in Figure 11.

![Figure 10](image-url)  
**Figure 10.** Network structure of the teacher and student models. (a) Resnet18 network structure (teacher model) and (b) Shuffv2_x0_5 network structure (student model).

4.3. Analysis of the Impact of Preprocessing on the Algorithm

This section verifies the effectiveness of the adaptive threshold segmentation method on the entire algorithm process. The segmentation scheme is based on the statistical characteristics of the photoelectric target echo light intensity. This section compares the number of candidate regions extracted by the threshold segmentation method and analyzes the effectiveness of the threshold segmentation method on the entire algorithm process. Due to the poor effect of some traditional threshold methods (see Section 4.1), this section only compares the screening detection area considering the Otsu threshold segmentation method. The candidate areas of 100 groups of active and passive images collected from outdoor scenes were selected, including 123 real targets. The shape characteristics of the candidate areas were screened consistently during the experiment, and only the threshold segmentation method was compared. The corresponding experimental results are illustrated in Figure 11.

![Figure 11](image-url)  
**Figure 11.** Comparison of the number of candidate regions produced by different threshold segmentation methods: (a) active image, (b) passive image, (c) Otsu-based segmentation, (d) candidate area segmentation based on the Otsu threshold, (e) proposed threshold segmentation effect, (f) candidate area segmentation based on the proposed threshold.

The red areas in Figure 11d,f represent the candidate region detected by the Otsu and the proposed threshold segmentation methods, respectively. The experimental comparison demonstrates that the developed threshold segmentation method significantly reduced the candidate region detected by the Otsu method and accurately contains the real target area. Especially for dynamic scenes, the suggested threshold segmentation method effectively filters out the peoples’ or targets’ relative motion trajectory, preserving the photoelectric target’s characteristic information.

In photoelectric target recognition, the detection rate \( P_d \) and false alarm rate \( P_f \) are important indicators to evaluate detection performance. The detection rate \( P_d \) is given by

\[
P_d = \frac{N_d}{N_t}
\]

where \( N_d \) and \( N_t \) are the number of real targets detected and existing in the image, respectively. The false alarm rate \( P_f \) is given by

\[
P_f = \frac{N_f}{N_r}
\]

where \( N_f \) is the number of false targets detected in the image and \( N_r \) is the number of targets detected in the image. The number of candidate regions generated by the threshold segmentation method affects the detection rate, false alarm rate, and inference
time of the photoelectric target detection algorithm and will ultimately impact the cat’s eye
detection performance.

The experimental results presented in Table 3 reveal that the Otsu threshold segmentation
method produces many candidate regions, decreasing the algorithm detection rate and
increasing the false alarm rate. More importantly, Otsu’s method significantly prolongs the
algorithm’s inference time, which is not conducive to its real-time detection performance.
In contrast, the suggested adaptive threshold segmentation method based on the statistical
characteristics of photoelectric target echo light intensity achieves an appealing perfor-
mance in detection rate, false alarm rate, and inference time, verifying the effectiveness of
the threshold segmentation method for cat-eye target recognition and the importance of
the entire algorithm process.

Table 3. Effects of different threshold segmentation methods on detection rate, false alarm rate, and
average inference time.

| Method   | Number of Candidate Regions | Detection Rate (%) | False Alarm Rate (%) | Average Inference Time |
|----------|----------------------------|--------------------|----------------------|------------------------|
| Otsu     | 1243                       | 93.49%             | 9.4%                 | 0.1781                 |
| Ours     | 167                        | 96.74%             | 4.8%                 | 0.0822                 |

4.4. TensorRT Acceleration and Algorithm Detection Results

This paper uses TensorRT technology [26] to afford the detection process, achieving
real-time performance on NVIDIA Jetson Nano. TensorRT is more suitable for hardware
to accelerate the inference process. Furthermore, this paper completes the transformation
of ONNX and constructs the acceleration engine (hereafter referred to as ShuffEng) based
on knowledge distillation to obtain the optimal network model. Since the NVIDIA Jetson
Nano platform only supports F32 precision, the engine adopts the full precision mode
for inference. Figure 12 illustrates a comparative test of the forward inference between
ShuffNet and ShuffEng accelerated by cudnn in GPU mode on NVIDIA Jetson Nano.

![Figure 12](image_url)

Figure 12. Forward reasoning comparison test: (a) active image, (b) inference result of ShuffNet,
(c) inference result of ShuffEng.
From the test set, 500 target candidate areas were selected and entered into ShuffNet and ShuffEng for discrimination. The discriminant results were evaluated using the detection rate $P_d$ and the false alarm rate $P_f$, as shown in Table 4. The weight precision of ShuffNet and ShuffEng is 32 bits. The above experimental results reveal that the detection results of ShuffNet and ShuffEng inference have not changed.

| Model   | Detection Rate | False Alarm Rate |
|---------|----------------|-----------------|
| ShuffNet| 97.15%         | 4.87%           |
| ShuffEng| 97.15%         | 4.87%           |

The subsequent trial is a comparative experiment on the inference time of ShuffNet and ShuffEng, involving an inference on 60 active and passive images with a resolution of 640 × 480 pixels on NVIDIA Jeston Nano. Since this method exploits the image’s preprocessing phase to determine the candidate region, the number of candidate regions will affect the inference time. In the process of selecting images, 30 images are single photoelectric target images, and 30 are multi-photoelectric target images. Figure 13 depicts the inference time of the entire algorithm under different conditions, revealing that the inference time of the entire algorithm process is relatively stable. Table 5 reports the average time of 30 inferences for ShuffNet and ShuffEng.

![Figure 13. Comparison of running time between single-target area and multi-target area.](image)

| Model   | Single-Target Region Inference Time (s) | Multi-Target Region Inference Time (s) |
|---------|----------------------------------------|----------------------------------------|
| ShuffNet| 0.0788                                 | 0.1293                                 |
| ShuffEng| 0.0341                                 | 0.0436                                 |

Figure 13. Comparison of running time between single-target area and multi-target area. (A) ShuNet inference for multi-target regions, (B) ShuNet inference for single-target regions, (C) ShuEng inference for multi-target regions, and (D) ShuEng inference for single-target regions.

Table 5. Comparison of average inference time between single-target regions and multi-target regions.
The experimental comparison highlights that ShuffEng has the same weight accuracy as ShuffNet but presents more advantages in inference. The ShuffEng generated in this paper achieves a 2.31 times acceleration ratio in single-optical target detection inference and 2.96 times acceleration ratio in multi-optical target detection inference, significantly improving the model's inference speed on the embedded platform. Its running time can be controlled in the range of 0.03–0.05 s, and the detection rate can reach 29 frames per second, meeting the real-time processing requirement of at least 20 frames per second in the video detection process. Therefore, the proposed method has the potential to be used in practical applications.

5. Conclusions

This paper proposes and verifies a new method of photovoltaic target detection based on NVIDIA Jeston Nano. The method mainly includes image preprocessing, knowledge distillation, TensorRT acceleration, and lightweight network target discrimination. The experimental results demonstrate that the developed adaptive threshold segmentation method based on the statistical characteristics of optical target light intensity improves detection of the target region more effectively than the traditional threshold segmentation methods. The layer-to-layer guidance of knowledge distillation is used to improve the accuracy of the lightweight network and TensorRT technology is used to realize accelerated reasoning on the NVIDIA Jeston Nano hardware platform. The proposed method has higher accuracy and lower requirements on hardware computing power and can quickly identify multiple photovoltaic targets. The entire process has good real-time performance, making the lightweight network more valuable in deploying low computing power devices.

Author Contributions: Conceptualization, H.S. and H.G.; methodology, L.Z.; software, S.Z.; validation, L.Z. and S.Z.; formal analysis, S.Z.; investigation, S.Z.; resources, L.Z.; data curation, L.Z.; writing—original draft preparation, S.Z.; writing—review and editing, S.Z. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Informed Consent Statement: Informed consent was obtained from all subjects involved in the study.

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

References

1. Huayan, S.; Laixian, Z.; Yanzhong, Z.; Yonghui, Z. Progress of Free-Space Optical Communication Technology Based on Modulating Retro-Reflector. Laser Optoelectron. Proc. 2013, 50, 040004. [CrossRef]
2. Laixian, Z.; Hua-yan, S.; Guihua, F.; Yanzhong, Z.; Yonghui, Z. Progress in free space optical communication technology based on cat-eye modulating retro-reflector. Chin. J. Opt. Appl. Opt. 2013, 6, 681–691.
3. Mieremet, A.-L.; Schleijpen, R.-M.-A.; Pouchelle, P.-N. Modeling the detection of optical sights using retroreflection. Proc. SPIE 2008, 6950, 69500.
4. Auclair, M.; Sheng, Y.; Fortin, J. Identification of Targeting Optical Systems By Multiwavelength Retrorereflection. Opt. Eng. 2013, 52, 54301. [CrossRef]
5. Anna, G.; Goudail, F.; Dolfi, D. General state contrast imaging: An optimized polarimetric imaging modality insensitive to spatial intensity fluctuations. Opt. Soc. Am. A 2012, 29, 892–900. [CrossRef] [PubMed]
6. Krizhevsky, A.; Sutskever, I.; Hinton, G.E. ImageNet classification with deep convolutional neural networks. In Proceedings of the Neural Information Processing Systems Conference, Lake Tahoe, NV, USA, 3–6 December 2012; pp. 1097–1105.
7. Simonyan, K.; Zisserman, A. Very deep convolutional networks for large-scale image recognition. arXiv 2014, arXiv:1409.1556.
8. He, K.; Zhang, X.; Ren, S.; Sun, J. Deep Residual Learning for Image Recognition. In Proceedings of the Neural Information Processing Systems Conference, Barcelona, Spain, 5–10 December 2016; pp. 770–778.
9. Sun, D.; Wang, Y.; Ni, D.; Wang, T. AutoPath: Image-Specific Inference for 3D Segmentation. Front. Neurorobotics 2020, 14, 49. [CrossRef] [PubMed]
10. Shao, H.; Wang, C.; Fu, Z.; Liu, Z. High-Resolution Distributed Radiation Detector System Assisted by Intelligent Image Recognition. Front. Phys. 2021, 9, 693838. [CrossRef]
11. Ke, X. Research on Hidden Camera Detection and Recognition Method Based on Machine Vision; Huazhong University of Science and Technology: Wuhan, China, 2020.
12. Liu, C.; Zhao, C.M.; Zhang, H.Y.; Zhang, Z.; Zhai, Y.; Zhang, Y. Design of an Active Laser Mini-Camera Detection System using CNN. *IEEE Photonics J.* 2019, 11, 1. [CrossRef]
13. Huang, J.H.; Zhang, H.Y.; Wang, L.; Zhang, Z.; Zhao, C. Improved YOLOv3 Model for miniature camera detection. *Opt. Laser Technol.* 2021, 142, 107133. [CrossRef]
14. Yanzhong, Z.; Huayan, S.; Xiaqiong, Y.; Mengshan, F. Three-Dimensional Analytical Formula for Oblique and Off-Axis Gaussian Beams Propagating through a Cat-Eye Optical Lens. *Chin. Phys. Lett.* 2010, 27, 034101. [CrossRef]
15. Ma, Y.; Tang, J. Foundations of Deep Learning. In *Deep Learning on Graphs*; Cambridge University Press: Cambridge, UK, 2021; pp. 43–72. [CrossRef]
16. Hinton, G.; Vinyals, O.; Dean, J. Distilling the Knowledge in a Neural Network. arXiv.org. Available online: https://arxiv.org/abs/1503.02531 (accessed on 10 August 2022).
17. Li, Z.; Yu, L.; Lu, B.; Fei, L.; Yawei, W. Using TensorRT for deep learning and inference applications. *J. Appl. Opt.* 2020, 41, 337–341. [CrossRef]
18. Shafi, O.; Rai, C.; Sen, R.; Ananthanarayanan, G. Demystifying TensorRT: Characterizing Neural Network Inference Engine on Nvidia Edge Devices. In Proceedings of the 2021 IEEE International Symposium on Workload Characterization (IISWC), Storrs, CT, USA, 7–9 November 2021; pp. 226–237. [CrossRef]
19. Wang, X.; Jun, Z.; Wang, S. The Cat-Eye Effect Target Recognition Method Based on Visual Attention. *Chin. J. Electron.* 2019, 28, 1080–1086. [CrossRef]
20. Romero, A.; Ballas, N.; Kahou, S.E.; Chassang, A.; Gatta, C.; Bengio, Y. Fitnets: Hints for thin deep nets. *arXiv 2014*, arXiv:1412.6550.
21. Jeong, E.; Kim, J.; Tan, S.; Lee, J.; Ha, S. Deep Learning Inference Parallelization on Heterogeneous Processors with TensorRT. *IEEE Embed. Syst. Lett.* 2021, 14, 15–18. [CrossRef]
22. Ningning, M.A.; Xiangyu, Z.; Haitao, Z.; Sun, J. ShuffleNet V2: Practical guidelines for efficient CNN architecture design. In Proceedings of the Computer Vision-ECCV 2018, Munich, Germany, 8–14 September 2018; pp. 116–131. [CrossRef]
23. Iandola, F.N.; Song, H.; Moskewicz, M.W.; Ashraf, K.; Dally, W.J.; Keutzer, K. SqueezeNet: AlexNet-level accuracy with 50x fewer parameters and<0.5 MB model size. *arXiv 2022*, arXiv:1602.07360.
24. Han, K.; Wang, Y.; Tian, Q.; Guo, J.; Xu, C.; Xu, C. Ghostnet: More features from cheap operations. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Seattle, WA, USA, 14–19 June 2020; pp. 1580–1589.
25. Yang, L.; Jiang, H.; Cai, R.; Wang, Y.; Song, S.; Huang, G.; Tian, Q. Condensenet v2: Sparse feature reactivation for deep networks. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Nashville, TN, USA, 20–25 June 2021; pp. 3569–3578.
26. Jeong, E.; Kim, J.; Ha, S. TensorRT-based Framework and Optimization Methodology for Deep Learning Inference on Jetson Boards. *Free Access 2022*. [CrossRef]