

Thanks to the advice of the reviewer. In order to make this paper more perfect, the following amendments are made to the questions raised by the reviewing teacher:

Q1: The result in line 17 of the article abstract " The detection rate of this method can reach 29 frames" is not reflected in the later text.

A1: Add the following to the conclusion: Its running time can be controlled in the range of 0.03-0.05s, 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.

Q2: The contents of Equation 10 are not explained.

A2: Equation 10 is a supplement to Equation 9 $\hat{q}_t$, and is explained in the paper. As follows: Equation 10 is a supplement to Equation 9.

Q3: In the title of the table in line 476 of the article, Table 4 should be Table 5.

A3: In the title of the table in line 476 of the article, Table 4 has been replaced by table 5. As follows:

Table 5. Comparison of average inference time between single-target regions and multi-target regions.

	Single-target Region	Multi-target Region
	Inference Time (/s)	Inference Time (/s)
ShuffNet	0.0788	0.1293
ShuffEng	0.0341	0.0436

Q4: The practical application value of this study is duly added.

A4: Add the following to the conclusion: The proposed method has higher accuracy and lower requirements on hardware computing power and can quickly identify multiple photoelectric targets. The entire process has good real-time performance, making the lightweight network more valuable in deploying low computing power devices.

Q5: There are grammatical and editorial issues in this paper. Many of them could cause difficulties to the readers.

A5: I have carefully reviewed all the information in this article. Many of the errors in this article have been corrected to ensure a better reading experience for the reader.

For example, the first paragraph of the original text: Pinhole cameras, miniature cameras, and other photoelectric equipment popularity bring a severe risk of information leakage, information leakage not only to the society and business brings huge property losses but also severe threats to the personal safety of individuals, for this phenomenon can be used the camera existing cat's eye effect [1] for detection, that is the incident light irradiation camera after the original way back and reflected light intensity than diffuse reflected light intensity 2-4 orders of magnitude higher. The spot generated by 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 does not rely on the start of the camera and the transmission of wireless signals, only on the physical characteristics of the camera itself, in this field the use of laser-active imaging for detection [2] is one of the most effective ways, mainly used in the detection of conventional cameras.

The first paragraph of the article after revision: Pinhole cameras, miniature cameras, and other photoelectric equipment impose a severe information leakage risk, which may impact society and business, bringing substantial property losses and severe threats to personal safety. Therefore, the camera's cat eye effect [1] is used for detection, utilizing the incident light irradiation camera after

the original way back and the reflected light intensity of the diffuse reflected light, which has a 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.

Q6: The chart format of this paper is not standardized. Some figures should not have solid lines on the outer ring; The typography of the figures is not aesthetically pleasing; The formatting of the tables in the article should be consistent.

A6: The chart format for this article has been standardized. The solid lines of some of the outer rings of the picture have been removed and the font has been unified.

Q7: In line 476, the table's serial number is mislabeled; Table 4 should be Table 5.

A7: In the title of the table in line 476 of the article, Table 4 has been replaced by table 5. As follows:

Table 5. Comparison of average inference time between single-target regions and multi-target regions.

	Single-target Region Inference Time (/s)	Multi-target Region Inference Time (/s)
ShuffNet	0.0788	0.1293
ShuffEng	0.0341	0.0436

Q8: In line 227, 7(c) appears in the article, but there is no indication in Figure 7 which figure (c) is the one.

A8: The content in the diagram is accurately labeled. 7(c) has been replaced by 6(c).

Q9: The author is suggested to supplement the significance of this research in practical application in the abstract and conclusion.

A9: Add the following to the conclusion: The experimental results demonstrate that the developed adaptive threshold segmentation method based on the statistical characteristics of optical target light intensity improves detecting the target region more effectively than the other 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 Jetson Nano hardware platform. The proposed method has higher accuracy and lower requirements on hardware computing power and can quickly identify multiple photoelectric targets. The entire process has good real-time performance, making the lightweight network more valuable in deploying low computing power devices.

Q10: Line 88 mentions that this method ensures the smooth acquisition of active and passive images. Please give a more specific explanation.

A10: Add the following to section 2.1: During image acquisition, the laser and detector are controlled by synchronous triggers. The detector acquires images every time the laser changes its operating state, and the detector's operating frequency is twice the laser's, ensuring smooth acquisition of active and passive images.

Q11: There is a word case error on line 315.

A11: The error has been modified. (, The output layer) to (, the output layer).

Q12: It is suggested that the author make a specific explanation to formula 10.

A12: Equation 10 is a supplement to Equation 9 $\hat{q}_l$ and is explained in the paper. As follows: Equation 10 is a supplement to Equation 9.

Q13: The author is requested to make a more detailed discussion on the results of the comparative test according to Fig. 9, so as to illustrate that the method proposed in this paper has certain advantages.

A13: Add the following: 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.