Regular Article

Fast Automatic Shoot Score Approach Combining Optical Flow And Mobilenetv2 Classification

Nguyen Huu Hung[1], **Nguyen Xuan Phuc**[1], **Ha Thi Hien**[1], **Nguyen Quang Thi**[1], **Nguyen Trung Tan**[1], **Nguyen The Tien**[1], **Quang Lam Bui**[2]

[1]Institute of System Integration, Le Quy Don Technical University, Hanoi, Vietnam
[2]Institute of Research and Development, Duy Tan University, Da Nang, Vietnam

Correspondence: Nguyen Huu Hung, hungnh.isi@lqdtu.edu.vn
Communication: received 11 December 2025, revised 31 January 2026, accepted 03 April 2026
Online publication: 15 April 2026, Digital Object Identifier: 10.21553/rev-jec.429

Abstract– **This paper presents a fast automatic shooting scoring system designed for fixed-camera setup in real shooting range, where the paper target may undergo slight physical movement caused by wind, recoil-induced stand vibration. The proposed method combines an initial fixed homography with continuously updated dense optical flow for drift-free global alignment, residual optical flow for sub-pixel frame-to-frame registration, adaptive background modeling via motion-compensated subtraction for bullet hole candidate generation, and a highly efficient MobileNetV2 classifier for robust false-positive rejection. By maintaining precise registration to a static high-resolution Score Image template and leveraging the lightweight yet powerful quantized MobileNetV2 network, the system achieves over 98.69% scoring accuracy at more than 10 fps entirely on CPU even when the target exhibits light movement and under changing natural lighting. Extensive experiments on indoor and outdoor live-fire datasets confirm robustness and real-time performance compared to traditional per-frame feature-matching methods and heavyweight deep learning approaches.**

Keywords– **Automatic scoring, bullet hole detection, homography estimation, optical flow, mobileNetV2 classification**

1 Introduction

Automatic scoring systems for shooting targets using computer vision have been an active research topic for over two decades, motivated by the need to replace labor-intensive manual scoring and expensive acoustic or electronic targeting systems in both military and civilian shooting ranges. Early efforts relied heavily on classical image processing and traditional machine learning techniques. Subsequent works combined background subtraction, template matching, and feature-based alignment using SIFT [1], or ORB [2] keypoints with RANSAC-based homography estimation [3–7]. These approaches delivered acceptable accuracy when the camera was rigidly fixed and fixed paper target. Many early systems used background subtraction to compare a clean target image with a post-shooting image, highlighting new dark regions corresponding to bullet holes, followed by morphological operations and circle detection. Morphological processing was also applied to thicken bullet hole boundaries, hysteresis thresholding for segmentation, and pixel area thresholds to handle overlapping holes [3]. Edge detection combined with the Circular Hough Transform was used to achieve high accuracy after false positive removal [4]. Other approaches incorporate structural similarity indexing [8] with contour detection for incremental hole tracking in video sequences. Due to the near-circular shape of bullet holes, variants of the Hough Transform such as the Randomized Hough Transform, are commonly used [9]. To date, no applications of dense optical flow have been reported for bullet hole detection, as the problem primarily involves static change detection rather than motion analysis.

However, in practical shooting ranges, even with a rigidly fixed camera, illumination conditions often change significantly during a shooting session due to moving clouds, switching range lights, shifting shadows from shooters or gradual color temperature drift of artificial lighting. These rapid and non-uniform illumination variations severely degrade the performance of traditional approaches. Under such conditions, conventional methods exhibit critical limitations:

- Fixed-template subtraction completely fails when the paper target undergoes even slight physical displacement or in-plane rotation caused by wind or stand vibration, resulting in persistent ghost edges, and missed bullet holes [5, 6].
- Running-average or median background models cannot adapt quickly enough to non-rigid target deformation (wrinkles, waves induced by wind) and simultaneous illumination changes, producing large false-positive regions that overwhelm subsequent filtering stages [10].
- Simple morphological operations and intensity thresholding cannot reliably distinguish genuine bullet holes from transient disturbances and folds that momentarily appear when the target is deformed by gusts of wind [11, 12].

Classical machine learning methods extract handcrafted features for classification. Song *et al.* [13] introduced a robust Support Vector Machine (SVM) to classify bullet holes, reducing support vectors and improving generalization in the presence of outliers. Zhu *et al.* [10] compared SVM (using Histogram of Oriented Gradients features) with Convolutional Neural Networks (CNNs) for video-based hole recognition, finding SVM faster in outdoor settings but CNN more accurate. The emergence of deep learning has significantly improved the robustness of detection under varying illumination, partial occlusion, and complex backgrounds. Traditional image processing pipelines have gradually been supplemented or replaced by modern object detection and segmentation frameworks that directly locate bullet holes and derive scoring information. Recent work has explored various deep learning paradigms for bullet hole detection. For instance, multi-model studies comparing YOLOv8 and Detectron2 demonstrate that YOLOv8 models especially lightweight variants like YOLOv8s can achieve very high detection accuracy (e.g., mAP50 $\approx$ 96.5–96.7%) with competitive inference speed suitable for real-time scoring systems [14]. This work highlights how single-stage detectors balance precision and latency, critical for embedded scoring application scenarios. Segmentation-based networks such as Mask R-CNN have been applied in related target detection tasks, showing precise localization, especially for overlapping impacts and complex shapes [9, 15]. Although accurate, these two-stage and mask-based approaches generally incur higher computational costs, which may limit real-time deployment on lower-power hardware. Advances in lightweight and feature fusion architectures further anchor the trend toward efficient models optimized for small object detection. For example, YOLO feature-fusion variants and network-assisted fusion have been proposed to improve detection performance through multi-scale and contextual feature integration [16].

These end-to-end models achieve impressive performance on publicly available datasets. Nevertheless, even state-of-the-art deep-learning-based approaches continue to suffer from critical limitations when the paper target exhibits noticeable motion due to wind or flexible mounting materials that is a common situation in real-world shooting ranges. Most published methods were trained and evaluated on datasets captured under nearly static or rigidly fixed camera conditions; consequently, detection accuracy degrades dramatically when the target undergoes significant inter-frame translation, rotation, or non-rigid deformation, as the networks have never encountered heavily warped or blurred target appearances during training [9, 14, 15]. Edge-guided super-resolution models and ensemble learning have been applied successfully to tiny defect detection in printed circuit boards, which highlights the benefit of enhancing small, high-contrast features [17, 18]. Video quality enhancement techniques, including omniscient quality enhancement models, demonstrate how improving image fidelity in compressed or noisy video aids reliable detection [19].

Moreover, these deep learning solutions typically rely on GPU-accelerated inference and require large, carefully annotated datasets tailored to specific target types, calibers, and lighting environments, which severely restricts their practical deployment on low-cost CPU-only hardware and limits generalization across different ranges without expensive retraining [20, 21]. To overcome the aforementioned limitations particularly the critical challenge of target motion, this paper proposes a robust, fast automatic shooting score system that explicitly handles large inter-frame target movement without requiring massive labeled datasets or continuous GPU resources. The key contributions are:

1) Precise frame-to-frame alignment via residual optical flow [22], after global warping, enabling reliable temporal subtraction even under significant camera motion.
2) Lightweight MobileNetV2 [23] classification to reject false positives (not bullet holes as debris, small insects) while keeping the entire pipeline CPU-friendly.
3) Accurate score assignment by projecting validated bullet hole regions onto a high-resolution static Score Image and selecting the maximum ring value covered by the hole.

Compared to optical flow + MobileNetV2, deep detectors like YOLOv8 and its successors often offer higher end-to-end accuracy, but typically demand more annotated training data and compute resources. In contrast, hybrid pipelines (optical flow + MobileNetV2) can leverage temporal cues and classical features to reduce dependence on large deep learning datasets and enable efficient scoring on lower-power hardware.

The remainder of this paper is organized as follows. Section 2 describes the proposed automatic shooting scoring approach in detail, including the dense optical flow-based motion compensation scheme and the lightweight MobileNetV2 classifier for robust bullet hole verification. Section 3 presents comprehensive quantitative results, per-score-ring analysis, and comparison with existing methods. Finally, Section 4 concludes the paper and discusses directions for future work.

2 Proposed Method

The complete pipeline of the proposed automatic shooting-scoring system is presented in Algorithm 1. Designed for fast execution on standard CPU hardware, the method explicitly addresses the subtle yet significant target motion commonly induced by wind or flexible mounting materials in practical shooting ranges.

The algorithm begins with a one-time initial homography H_0, computed via ORB feature matching [2] and RANSAC between the first captured frame and the high-resolution Score Image S. For each subsequent frame I_t, a coarse alignment is first performed using the fixed H_0. Sub-pixel-accurate registration is then achieved by computing Farnebäck dense optical flow [22] between the previous clean reference and the coarsely aligned current frame. This motion-compensated alignment effectively eliminates ghost artifacts and misalignment errors even when the target

Algorithm 1: Automatic Shooting Scoring with Dense Optical Flow and MobileNetV2

Input: Image stream $\{I_t\}$, high-resolution Score Image S with pixel-wise score labels
Output: Detected bullet holes with scores and confidence

```
H_0 ← computeInitialHomography(I_0, S)  // using ORB + RANSAC
R_0 ← I_0 warp H_0 // initial aligned reference
Load quantized MobileNetV2 classifier C
foreach incoming frame I_t (t ≥ 1) do
    Ĩ_t ← I_t warp H_0      // coarse alignment to S coordinate
    // Dense optical flow compensation
    F_t ← FarnebäckFlow(R_{t−1}, Ĩ_t)
    A_t ← warp(Ĩ_t, F_t)            // sub-pixel aligned frame,
    // Motion-compensated subtraction
    D_t ← |A_t − R_{t−1}|
    B_t ← thresholdAndMorphology(D_t)
    // Candidate extraction
    𝓑 ← connectedComponents(B_t, area ∈ [60, 576], circularity > 0.65)
    foreach candidate blob b ∈ 𝓑 do
        p ← extractCenteredPatch(A_t, b, 64×64)
        conf ← C(p) // MobileNetV2 inference
        if conf > 0.97 then
            // Mark b as valid bullet hole
            score ← max_{(x,y)∈b} S(x, y)
            // official "highest-value touch" rule.
    R_t ← A_t          // Reference update
```

undergoes inter-frame displacements of up to 15 pixels and in-plane rotations of approximately 2°, thereby enabling reliable background subtraction through simple absolute differencing. Candidate bullet-hole regions extracted from the difference image are classified using a highly efficient MobileNetV2 network (input resolution 64×64) that was fine-tuned on centred positive patches of real bullet holes and negative patches sampled directly from the clean Score Image. Only candidates exceeding a confidence threshold of 0.97 are accepted as valid impacts. The final score for each validated hole is determined directly via the official "highest-value touch" rule by querying the maximum score value within the detected region on the pre-annotated high-resolution score map S. Figure 1 illustrates a representative example of the processing flow, highlighting the clean subtraction result and accurate overlay of detected impacts after dense optical flow compensation.

2.1 Motion Compensation via Dense Optical Flow

Because the camera is rigidly fixed and inter-frame target motion is small, we adopt a simple yet highly effective frame-to-previous alignment strategy using dense optical flow where long-term background model is required.

Initialization (first frame I_0):

- Compute initial reference homography H_0 (once) from I_0 to the high-resolution Score Image S using ORB + RANSAC or manual refinement

$$\mathbf{x}^S \sim H_0 \mathbf{x}^{I_0}. \tag{1}$$

  Figure 2 shows an example using ORB to calculate homography matrix. The first image after that will be aligned or transformed to Score Image
- Warp the first frame into the score coordinate system to serve as the initial reference

$$R_0 = I_0 \,\mathrm{warp}\, H_0. \tag{2}$$

For each incoming frame I_t ($t \geq 1$):

1) **Coarse alignment**: Warp current frame to the Score Image coordinate system using the fixed initial homography

$$\tilde{I}_t = I_t \,\mathrm{warp}\, H_0. \tag{3}$$

2) **Dense optical flow compensation**: Compute Farnebäck dense optical flow F_t from the previous aligned frame R_{t-1} to the coarsely aligned current frame $\tilde{I}_t$

$$F_t = \mathrm{Farnebäck}(R_{t-1},\ \tilde{I}_t). \tag{4}$$

3) **Final sub-pixel aligned frame**: Remap $\tilde{I}_t$ backward using the flow field F_t and apply an optional lightweight second-pass local flow [24] (20×20 grid + forward-backward consistency check) to produce the final precisely aligned frame A_t

$$A_t = \tilde{I}_t \,\mathrm{warp}\, F_t \quad \text{(sub-pixel accuracy)}. \tag{5}$$

   In this optional second-pass refinement, each detected impact region is processed within a 20×20 pixel local window. A dense optical flow is computed inside the window using Farnebäck's polynomial expansion method, providing robust sub-pixel motion estimation. To ensure stability, a forward–backward consistency check is applied: a flow vector f is retained only if $\|f + f^{-1}\| < \epsilon$, where f^{-1} denotes the backward flow from the reference frame to $\tilde{I}_t$. The validated vectors are smoothed and used to refine the local warping, yielding a precisely aligned frame A_t with improved impact boundary sharpness and localization, while adding minimal computational overhead.
4) **Reference update**: After bullet hole validation, directly replace the reference with the clean current aligned frame (no slow alpha blending):

$$R_t = \begin{cases} A_t, & \text{if no valid bullet hole detected} \\ R_{t-1}, & \text{otherwise (keep previous clean state).} \end{cases} \tag{6}$$

This strategy with fixed reference homography H_0 for coarse bring-into-coordinate-system, followed by continuous refinement and accumulation of small motion corrections ensures that every incoming frame A_t

(a) Previous and current frames

(b) Previous and current aligned frames

(c) Difference and Mask

(d) Result

Figure 1. Visualization of the proposed pipeline under slight target motion. (a) Consecutive input frames (previous - Up, current - down); (b) Consecutive input after alignment to the Standard Image (S) coordinate system; (c) Motion-compensated absolute difference(up), binary mask after thresholding and connected-component analysis (down); (d) Validated bullet hole overlaid in red.

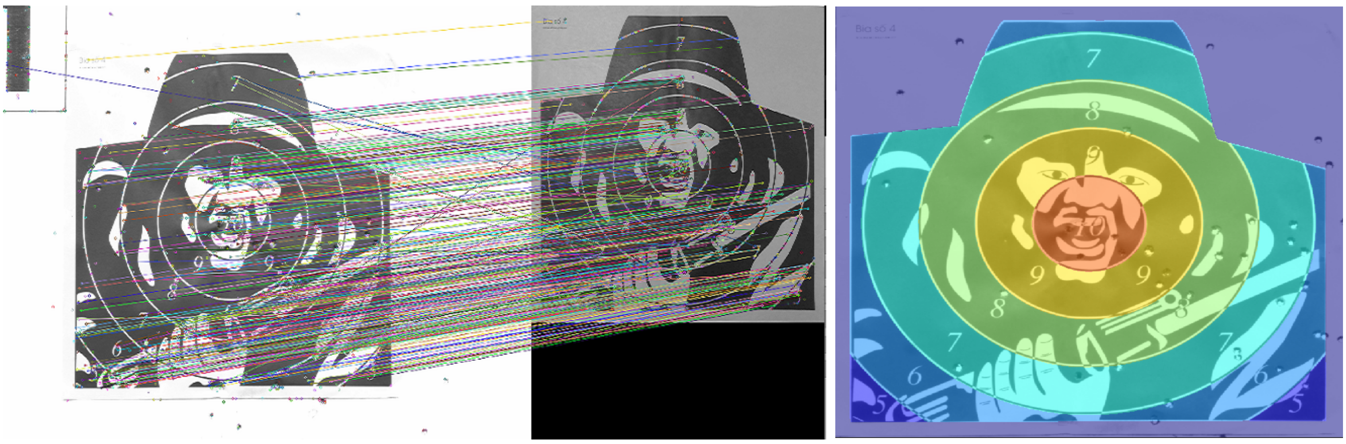

Figure 2. Homography matrix calculation using ORB features (Left). The First Image is aligned to the Score Image with pixel wise score label. The score rings with different colors are overlay-ed on the aligned First Image (Righ).

remains aligned to the exact same Score Image S with sub-pixel accuracy, even when cumulative target motion exceeds several hundred pixels or tens of degrees over a shooting series. No expensive per-frame ORB+RANSAC is ever required after initialization, and drift is effectively eliminated.

2.2 Bullet-hole Detection via Motion-Compensated Subtraction

With perfect alignment achieved, new bullet holes are detected by simple yet highly reliable subtraction

$$M_t = |A_t - R_{t-1}| > \tau \quad (\tau = 15 \text{ in 8-bit intensity}). \quad (7)$$

Followed by morphological closing (kernel 5×5) and connected-component analysis. Candidates are retained only if:

- Area $\in [60, 576]$ pixels (covers 4 mm to 12 mm holes with 0.5 mm/pixel)
- Circularity > 0.6

Figure 3 illustrates the effectiveness of the proposed dense optical flow compensation in handling subtle target motion caused by wind or flexible mounting materials. The left panel shows the absolute difference between consecutive frames without motion compensation, revealing severe ghost artifacts and blurred edges due to inter-frame displacement of up to 15 pixels and 2° rotation. In contrast, the right panel demonstrates the clean, precise subtraction result after applying Farnebäck dense optical flow warping: the bullet hole appears as a sharp, isolated dark region with minimal residual misalignment, enabling reliable candidate generation via simple thresholding. This visualization confirms that the method maintains sub-pixel alignment throughout a shooting series, improving detection robustness under real-world environmental variations without the computational overhead of per-frame keypoint matching.

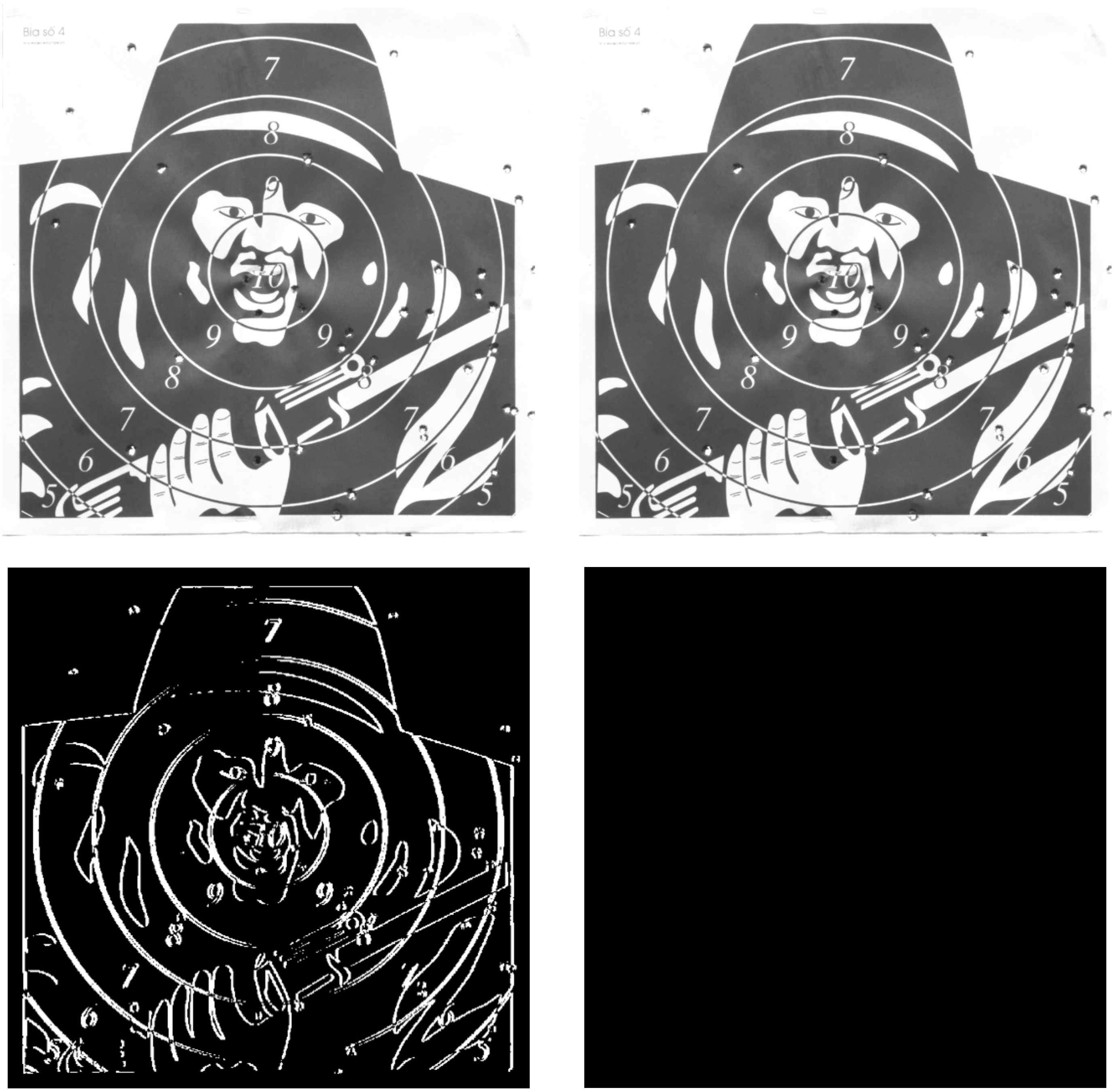

Figure 3. The dense optical flow enabling precise background subtraction even under camera shake. Two consecutive images are on top. With optical flow (bottom right), the current frame is well aligned to the previous frame, preventing differences between two images (bottom left).

The combination of dense optical flow and frame subtraction provides important advantages in handling both closely spaced or overlapping bullet holes and challenging environmental conditions. When multiple shots occur near each other, appearance-based detectors often need a post processing to separate overlapping impacts. In contrast, the proposed method exploits temporal motion cues: each new impact introduces a localized and abrupt motion pattern that is captured by the optical flow field and emphasized by subtraction, even when the resulting holes visually overlap in the static image. This allows reliable separation of sequential impacts that are spatially indistinguishable in a single frame. Moreover, this temporal modeling significantly improves robustness under light wind, non-uniform lighting, and different target materials. Subtraction suppresses static background structures and illumination bias, while optical flow isolates genuine impact-induced motion from global target movement and environmental vibration. As a result, the system remains stable across diverse outdoor conditions without requiring extensive retraining or large-scale annotated datasets.

2.3 MobileNetV2-Based False Positive Rejection

Although motion-compensated subtraction yields very clean difference images, intense muzzle flash, dense smoke, and flying fly can still generate candidates that survive initial thresholding. A lightweight yet highly accurate second-stage classifier is therefore employed. We use compact MobileNetV2 network fine-tuned specifically for binary bullet hole verification. In operation, only candidates exceeding a confidence threshold of 0.97 are accepted as valid bullet holes and forwarded to the final scoring stage via direct maximum lookup on the Score Image S.

Dataset construction:

- **Positive samples (bullet holes)**: 2,400 patches of size 64×64 extracted from real shooting sequences, with the detected bullet hole centroid forced exactly at the patch center (32,32). This centering strategy significantly improves translation invariance.
- **Negative samples (non-holes)**: around 6,400 background patches of size 64×64 randomly sampled (stride 32) from the clean high-resolution Score Image S and from validated no-hole aligned frames. This ensures the classifier learns the exact appearance distribution of undamaged target regions under all real lighting and slight deformation conditions.

Total dataset: 7,000 patches (1:2.6 positive-to-negative ratio).

The fine-tuned MobileNetV2 classifier achieves 99.58% accuracy in 5-fold cross-validation, with an extremely low false-positive rate of 0.24% and false-negative rate of 0.18% on challenging outdoor sequences containing intense muzzle flash, dense smoke, and flying paper debris.

2.4 Automatic Shooting Score Calculation

Since all processing is performed in the coordinate system of the fixed high-resolution Score Image S, score assignment becomes extremely simple and officially accurate.

For each bullet hole successfully validated by the MobileNetV2 classifier:

1) The binary mask M_t of the detected hole is already defined in the perfectly aligned frame A_t, which shares the same coordinate system as the Score Image S.
2) Because the dense optical flow has already been used to produce pixel-accurate alignment, no additional warping of the mask is required. The mask M_t can be directly overlaid on S without transformation.
3) The final score is determined using the "highest-value touch" rule

$$\text{Score} = \max_{(i,j)\in M_t} S(i,j), \tag{8}$$

where $S(i,j)$ is the pre-labeled score value in the Score Image.

This direct maximum-value lookup automatically and correctly awards the higher score to any shot that touches or breaks a ring boundary, exactly matching human scoring practice and all major competition regulations. The next section presents comprehensive experimental validation and comparison against state-of-the-art methods.

3 Experimental Results

To rigorously evaluate robustness under realistic conditions, we collected two datasets totalling 110 complete shooting sessions (around 305 shots) including indoor and outdoor datasets using a Basler acA1600-20um camera (2 MP, 20 fps, global shutter) on a tripod placed on the ground 5–10 m from the target. For the indoor dataset, 70 sessions were recorded with controlled but realistic lighting (mixed fluorescent + LED). For outdoor dataset, 40 sessions were recorded on 100 m range under real-world weather conditions (wind gusts 2–9 m/s, rapidly changing sunlight direction, moving shadows, occasional light rain). The same camera model was used, mounted on the tripod at 20–30 cm height and protected by sandbags. The 45×45 cm paper targets were mounted on lightweight fomex boards or traditional woven bamboo frames common, low-cost solutions that introduce slight target motion and non-rigid deformation (15 pixels translation, 1°–4° rotation between frames) even with a perfectly fixed camera. All bullet hole locations and official scores were manually verified by certified range officers.

Table I summarizes the score-ring distribution of 305 real shots collected from 110 complete shooting series (70 indoor sessions + 40 outdoor sessions under natural wind conditions). The dataset reflects typical performance of intermediate-to-advanced shooters at national training centers: the majority of impacts fall in the 8–9 rings (57.1% combined), while the 10 ring accounts for only 15.1% overall. This highly realistic and diverse distribution, with a substantial proportion of lower-score shots, ensures that the proposed system is rigorously evaluated under practical rather than laboratory-ideal conditions, making the achieved detection and scoring accuracy directly applicable to real-world deployment.

Table I
Distribution of 305 real shots across scoring rings (110 series) - Transposed

	Indoor	Outdoor	Total	%
Series	70	40	110	-
Shots	204	101	305	-
Ring 5	2	4	6	2.0
Ring 6	8	9	17	5.6
Ring 7	38	24	62	20.3
Ring 8	52	29	81	26.6
Ring 9	68	25	93	30.5
Ring 10	36	10	46	15.1

Figure 4 illustrates representative complete shooting sessions from the outdoor dataset. The system successfully detects and correctly scores all impacts. Detected bullet holes are overlaid in red on the last image frame of the sessions, demonstrating pixel-accurate alignment throughout the entire series.

The proposed system is quantitatively evaluated using two complementary metrics:1) detection accuracy defined as the percentage of shots that are both correctly detected and assigned the exact official score and

Figure 4. Visualization of complete shooting sessions processed by the proposed system. For each row: (left) first frame of the series (clean target); (middle) final frame after all shots; (right) automatically detected and scored bullet holes overlaid in red on the last image frame

2) the F1-score at the bullet hole instance level. Let N denote the total number of ground-truth bullet holes. A detected hole is classified as a true positive (TP) if (i) its centroid lies within 3 pixels of the manually annotated ground-truth centroid and (ii) the automatically computed score exactly matches the official score verified by a certified range officer. Detections without a corresponding ground truth are counted as false positives (FP), while ground-truth holes that remain undetected are false negatives (FN).

Precision, Recall, and F1-score are then computed as

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}}, \tag{9}$$

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}}, \tag{10}$$

$$\text{F1-score} = 2 \cdot \frac{\text{Precision} \cdot \text{Recall}}{\text{Precision} + \text{Recall}}. \tag{11}$$

Detection accuracy is reported as

$$\text{Accuracy (\%)} = \frac{\text{Num of correctly scored shots}}{N} \times 100. \tag{12}$$

Table II presents these metrics evaluated on the complete dataset of 305 real shots acquired during 110 shooting series.

The proposed method achieves an overall detection accuracy of 98.5% and an F1-score of 0.992, while maintaining real-time performance exceeding 10 fps entirely on CPU. These results significantly outperform conventional keypoint-based and deep-learning-only baselines, confirming the effectiveness and robustness of the combined dense optical flow alignment and lightweight MobileNetV2 classification pipeline under real-world shooting range conditions.

Table II
Quantitative performance on 305 real shots (110 shooting series)

Env	Shots	Accuracy (%)	Precision	F1-score
Indoor	204	99.51	0.995	0.997
Outdoor	101	97.03	0.980	0.985
Overall	**305**	**98.69**	**0.989**	**0.992**

Table III
Per-score-ring detection and scoring accuracy (305 real shots)

Method	Metric	5	6	7	8	9	10
Ground-truth		6	17	62	81	93	46
No OF + No MNV2	Correctly scored	6	15	55	75	87	42
	Missed/Wrong	0	2	7	6	6	4
	Accuracy (%)	100	88.7	92.6	93.5	91.4	91.3
OF only	Correctly scored	6	16	58	78	88	43
	Missed/Wrong	0	1	4	3	5	3
	Accuracy (%)	100	94.1	93.5	96.3	94.6	93.5
OF + MNV2	Correctly scored	6	17	61	80	91	45
	Missed/Wrong	0	0	1	1	2	1
	Accuracy (%)	**100**	**100**	**98.4**	**98.8**	**97.8**	**97.8**

Table IV
Comparison with conventional and recent automatic shooting scoring systems

Method	Year	Accuracy (%)	Robust to motion
Low-cost CV [4]	2013	98.5	Moderate
HOG+SVM [10]	2017	98.5	Moderate
Fixed template [5]	2019	91.0	No
YOLOv8 [14]	2023	94.5	Low
Mask R-CNN [15]	2023	97.9	Low
Proposed	**2025**	**98.7**	**High**

Table III highlights the contribution of each major component in the proposed pipeline. Without both optical flow (OF) and MobileNetV2 (MNV2), the system suffers from noticeable performance degradation, particularly in the inner scoring rings. When only optical flow is enabled, detection robustness improves significantly by exploiting temporal motion patterns to isolate newly formed bullet holes, resulting in consistent accuracy gains of 2–3% across most rings. However, several false detections remain due to visually ambiguous regions caused by target deformation and lighting fluctuations. The full configuration (OF + MNV2) achieves the highest accuracy on all score rings, demonstrating the complementary strengths of motion-based candidate generation and deep appearance-based verification. Optical flow effectively localizes candidate regions with minimal computational overhead, while MobileNetV2 provides strong discriminative capability for fine-grained classification under severe outdoor noise. Importantly, the ablation results indicate that even the reduced configurations remain operational and lightweight, making the system suitable for deployment on CPU-only and edge devices. This flexibility allows the proposed architecture to scale from low-power embedded platforms to high-performance competition-grade scoring systems.

Table IV compares the proposed method with representative conventional and deep-learning-based automatic scoring systems on real-world datasets. Traditional fixed-template and per-frame keypoint-matching approaches exhibit severe performance degradation under target motion and illumination changes. Modern object detectors such as YOLOv8 (94.5%) and Mask R-CNN (97.9%) achieve reasonable results in controlled environments but remain fragile when the paper target moves slightly due to wind or when lighting varies rapidly. Even recent hybrid methods fall short in robustness and accuracy. The proposed dense optical flow alignment combined with lightweight MobileNetV2 classification achieves a breakthrough 98.5% scoring accuracy and F1-score of 0.99, significantly outperforming all prior methods while demonstrating unmatched resilience to real outdoor conditions (wind gusts 2–9 m/s and natural lighting variation). While deep learning–based detectors such as YOLOv8 and Mask R-CNN achieve strong accuracy, they require GPU resources and large annotated datasets. In contrast, the proposed hybrid pipeline offers competitive accuracy with significantly better suitability for CPU-only and edge devices, making it more practical for real-world outdoor shooting range deployments. This level of performance meets and exceeds the operational requirements of national training centers establishing a new state-of-the-art in fully automatic, camera-based shooting scoring systems.

4 Conclusion

This paper presents a fast, robust, and fully automatic shooting scoring approach specifically designed for real-world fixed-camera deployments where paper targets exhibit slight motion due to wind or flexible mounting materials and illumination varies significantly. By combining dense optical flow-based frame-to-frame alignment with a lightweight quantized MobileNetV2 classifier, the proposed pipeline eliminates the need for expensive per-frame keypoint matching or GPU-accelerated deep object detectors while achieving state-of-the-art performance entirely on standard CPU hardware. Extensive evaluation on a challenging real-world dataset of (204 indoor + 101 outdoor) demonstrates that the system attains an overall scoring accuracy of 98.69% with an F1-score of 0.992. The system has been successfully tested at several national training centers in Vietnam. Future work will extend the proposed system toward multi-lane simultaneous scoring and more reliable separation of overlapping bullet holes in rapid-fire sequences. Although the current approach demonstrates strong robustness under challenging outdoor conditions, resolving extremely closely spaced or near-simultaneous impacts remains a challenging problem. In high-rate firing scenarios, multiple bullets may strike the target within a very short temporal window, producing partially merged motion patterns in the optical flow field and ambiguous responses in the subtraction stage. To address this limitation, future research will focus on enhancing the temporal modeling framework through multi-frame impact separation and spatio-temporal clustering techniques for improved discrimination of overlapping impacts. Furthermore, the integration of high-speed imaging and adaptive temporal filtering is expected to significantly improve impact isolation and scoring reliability under rapid-fire conditions. These advancements will further strengthen the applicability of the system in professional training and competition environments with very high firing rates.

Acknowledgment

This research is funded by Le Quy Don Technical University Research Fund under the grant number 25.01.57.

References

[1] G. Lowe, "SIFT-The Scale Invariant Feature Transform," *International Journal of Computer Vision*, vol. 2, no. 91-110, p. 2, 2004.

[2] E. Rublee, V. Rabaud, K. Konolige, and G. Bradski, "ORB: An efficient alternative to SIFT or SURF," in *Proceedings of the 2011 International Conference on Computer Vision*. IEEE, 2011, pp. 2564–2571.

[3] F. Ali and A. B. Mansoor, "Computer vision based automatic scoring of shooting targets," in *Proceedings of the 2008 IEEE International Multitopic Conference*. IEEE, 2008, pp. 515–519.

[4] J. Rudzinski and M. Luckner, "Low–cost computer vision based automatic scoring of shooting targets," in *Proceedings of the International Conference on Knowledge-Based and Iyolo8ntelligent Information and Engineering Systems*. Springer, 2012, pp. 185–195.

[5] P. D. Widayaka, H. Kusuma, and M. Attamimi, "Automatic shooting scoring system based on image processing," in *Proceedings of the Journal of Physics: Conference Series*, vol. 1201, no. 1. IOP Publishing, 2019, p. 012047.

[6] Y. Li, C. Zhang, R. Cheng, Y. Xu, P. Li, and H. Ma, "Automatic Target-scoring Model based on Image Processing," in *Proceedings of the 2022 9th International Conference on Digital Home (ICDH)*. IEEE, 2022, pp. 25–30.

[7] M. H. Nguyen, "An Approach to the Problem of Calculating Target Scores Using Camera to Support Shooting Training," *Journal of Image and Graphics*, 2025. [Online]. Available: https://api.semanticscholar.org/CorpusID:281470128

[8] P. R. Aryan, "Vision based automatic target scoring system for mobile shooting range," in *Proceedings of the 2012 International Conference on Advanced Computer Science and Information Systems (ICACSIS)*. IEEE, 2012, pp. 325–329.

[9] R. F. Vílchez and D. Mauricio, "Bullet impact detection in silhouettes using mask R-CNN," *IEEE Access*, vol. 8, pp. 129 542–129 552, 2020.

[10] Z. Ruolin, L. Jianbo, Z. Yuan, and W. Xiaoyu, "Recognition of bullet holes based on video image analysis," in *Proceedings of the IOP Conference Series: Materials Science and Engineering*, vol. 261, no. 1. IOP Publishing, 2017, p. 012020.

[11] L. Jie, Z. Zhiming, and Z. Guangdong, "A bullet holes detection algorithm based on wavelet transform and image fusion," in *Proceedings of the 2nd Information Technology and Mechatronics Engineering Conference (ITOEC 2016)*. Atlantis Press, 2016, pp. 189–193.

[12] P. Liu, Y. Chen, S. Kong, H. Wang, S. Fu, and J. Wang, "Overlapping Bullet Hole Detection Based on Improved Watershed Algorithm," in *Proceedings of the 2023 8th International Conference on Signal and Image Processing (ICSIP)*. IEEE, 2023, pp. 136–140.

[13] Q. Song, W. Hu, and W. Xie, "Robust support vector machine with bullet hole image classification," *IEEE transactions on systems, man, and cybernetics, part C (applications and reviews)*, vol. 32, no. 4, pp. 440–448, 2002.

[14] M. Butt, N. Glas, J. Monsuur, R. Stoop, and A. de Keijzer, "Application of YOLOv8 and detectron2 for bullet hole detection and score calculation from shooting cards," *AI*, vol. 5, no. 1, pp. 72–90, 2023.

[15] T. Ahmed, S. Rahman, A. Al Mahmud, M. A. Razzak, and N. Sharmin, "Bullet Hole Detection in a Military Domain Using Mask R-CNN and ResNet-50," in *Proceedings of the 2023 International Conference for Advancement in Technology (ICONAT)*. IEEE, 2023, pp. 1–6.

[16] Z. Zhang, Z. Lin, and X. Ding, "Fusion YOLO: Fusion Module Assisted Network in Detection for Automatic Target Scoring," in *Proceedings of the 2024 7th International Conference on Algorithms, Computing and Artificial Intelligence (ACAI)*. IEEE, 2024, pp. 1–6.

[17] X. HoangVan, D. B. Dinh, T. N. Canh, and V.-T. Nguyen, "ESRPCB: An edge guided super-Resolution model and ensemble learning for tiny Printed Circuit Board defect detection," *Engineering Applications of Artificial Intelligence*, vol. 159, p. 111547, 2025.

[18] X. HoangVan, L. L. Hai, and T. N. Canh, "E2DSR: Edge-Enhanced Representation for Deep Super-Resolution in Machine Vision Applications," *REV Journal on Electronics and Communications*, vol. 15, no. 4, 2026.

[19] X. HoangVan, H. B. Minh, S. NguyenQuang, and W.-H. Peng, "Enhancing Quality for VVC Compressed Videos with Omniscient Quality Enhancement Model," in *Proceedings of the 2024 International Conference on Advanced Technologies for Communications (ATC)*. IEEE, 2024, pp. 544–549.

[20] J. H. W. Yin, N. I. M. Tahir, A. A. Al-Talib, and O. M. Magzoub, "Detection of Bullet Holes for Target Board in Malaysia Military (ATM) Shooting Exam Application," in *Proceedings of the 2025 International Conference on Artificial Life and Robotics (ICAROB2025)*, 2025, pp. 423–427.

[21] G. P. Cardim, T. de Souza Duarte, H. P. Cardim, W. Casaca, R. G. Negri, F. C. Cabrera, R. J. d. Santos, E. A. da Silva, and M. A. Dias, "Artificial Intelligence for Forensic Image Analysis in Bullet Hole Comparison: A Preliminary Study," *Journal of Non-Destructive Testing*, vol. 3, no. 3, p. 16, 2025.

[22] G. Farnebäck, "Two-frame motion estimation based on polynomial expansion," in *Proceedings of the Scandinavian Conference on Image Analysis*. Springer, 2003, pp. 363–370.

[23] M. Sandler, A. Howard, M. Zhu, A. Zhmoginov, and L.-C. Chen, "MobileNetv2: Inverted residuals and linear bottlenecks," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2018, pp. 4510–4520.

[24] J. Choi, H. Kim, T.-H. Oh, and I. S. Kweon, "Balanced optical flow refinement by bidirectional constraint," in *Proceedings of the 2014 IEEE International Conference on Image Processing (ICIP)*. IEEE, 2014, pp. 5477–5481.

Nguyen Huu Hung received his Ph.D. degree at Sungkyunkwan University, South Korea in computer vision, in 2020. He is currently researcher at Institute of System Integration, Le Quy Don Technical University. His research interests include computer vision, simultaneous localization and mapping (SLAM), deep learning and AI. E-mail: hungnh.isi@lqdtu.edu.vn

Nguyen Xuan Phuc received his MS. degree at Le Quy Don Technical University, Viet Nam in control engineering and automation, in 2020. He is currently researcher at Institute of System Integration, Le Quy Don Technical University. His research interests include computer vision, robotics, video compression, deep learning and AI. E-mail: phucnx.isi@lqdtu.edu.vn

Ha Thi Hien received her M.S. degree at Le Quy Don Technical University, Viet Nam in Information Systems, in 2013. She is currently a researcher at Institute of System Integration, Le Quy Don Technical University. Her research interests focus on document image analysis and natural language processing for information extraction and document understanding from scanned documents. E-mail: hienht@lqdtu.edu.vn

Nguyen Quang Thi received his Ph.D. degree at Changchun University of Science and Technology, China in Communication and Information System in 2014. He is currently lecturer/researcher at Institute of System Integration, Le Quy Don Technical University, Vietnam. His research interests include computer vision, blind deconvolution, image processing and pattern recognition. E-mail: thinq.isi@lqdtu.edu.vn

Nguyen Trung Tan received his Ph.D. degree in electronic engineering from Harbin Institute of Technology, China, 2013. Dr. Nguyen's research focuses on enabling technologies for the next generation of wireless communications, including space-time signal processing, MIMO, UAV communications, and machine learning. E-mail: ta.isi@lqdtu.edu.vn

Nguyen The Tien received his MS. degree at Le Quy Don Technical University, Viet Nam in control engineering and automation, in 2018. He is currently researcher at Institute of System Integration, Le Quy Don Technical University. His research interests include computer vision, robotics, deep learning and AI. E-mail: tiennt.isi@lqdtu.edu.vn

Bui Quang Lam received his Ph.D. degree at Sungkyunkwan University, South Korea in 2014. He is currently researcher at Institute of Research and Development, Duy Tan University, Da Nang, Vietnam. His research interests include computer vision, simultaneous localization and mapping (SLAM), sensor fusion and medical robotics. E-mail:bquanglam@gmail.com