MDPI

Article

LDFE-SLAM: Light-Aware Deep Front-End for Robust Visual SLAM under Challenging Illumination

Cong Liu [1,2,*], You Wang [1], Weichao Luo [1] and Yanhong Peng [2,*]

[1] Peng Cheng Laboratory, Shenzhen, China
[2] College of Mechanical Engineering, Chongqing University of Technology, Chongqing 400054, China
* Correspondence: liucong@pcl.ac.cn (C.L.); yhpeng@nagoya-u.jp (Y.P.)

Abstract

Visual SLAM systems face significant performance degradation under dynamic lighting conditions, where traditional feature extraction methods suffer from reduced keypoint detection and unstable matching. This paper presents LDFE-SLAM, a novel visual SLAM framework that addresses illumination challenges through a Light-Aware Deep Front-End (LDFE) architecture. Our key insight is that low-light degradation in SLAM is fundamentally a geometric feature distribution problem rather than merely a visibility issue. The proposed system integrates three synergistic components: (1) an illumination-adaptive enhancement module based on EnlightenGAN with geometric consistency loss that restores gradient structures for downstream feature extraction, (2) SuperPoint-based deep feature detection that provides illumination-invariant keypoints, and (3) LightGlue attention-based matching that filters enhancement-induced noise while maintaining geometric consistency. Through systematic evaluation of five method configurations (M1–M5), we demonstrate that enhancement, deep features, and learned matching must be co-designed rather than independently optimized. Experiments on EuRoC and TUM sequences under synthetic illumination degradation show that LDFE-SLAM maintains stable localization accuracy (∼1.2m ATE) across all brightness levels, while baseline methods degrade significantly (up to 3.7m). Under Severe (25% brightness) conditions, our method achieves 62% tracking success rate compared to 12% for ORB-SLAM3, with keypoint detection remaining above the critical 100-point threshold even under Extreme degradation.

Keywords: visual SLAM; low-light environment; LightGlue; SuperPoint; illumination-adaptive; point-line fusion; deep feature matching; robust localization

Received:
Revised:
Accepted:
Published:

Citation: Liu, C.; Wang, Y.; Luo, W.; Peng, Y. LDFE-SLAM: Light-Aware Deep Front-End for Robust Visual SLAM. *Sensors* **2025**, *1*, 0. https://doi.org/

1. Introduction

Visual Simultaneous Localization and Mapping (SLAM) has become a fundamental technology for autonomous systems, enabling robots, drones, and augmented reality devices to navigate and understand their environments [1,2]. However, real-world deployment of visual SLAM faces a critical challenge: performance degradation under dynamic lighting conditions, including low-light environments, rapid illumination changes, and high dynamic range scenes [3,4].

The core issue in low-light visual SLAM is not merely insufficient brightness—it is the degradation of geometric feature distributions [5,6]. When illumination decreases, traditional feature detectors such as ORB experience a dramatic reduction in keypoint quantity (often 40–60% fewer points), while the remaining features exhibit poor spatial

distribution and reduced repeatability [7]. This geometric degradation leads to tracking failures, trajectory drift, and ultimately system breakdown.

Previous approaches to address illumination challenges in visual SLAM fall into two categories. The first category employs image enhancement as a preprocessing step, applying techniques such as histogram equalization, Retinex-based methods, or deep learning-based enhancement networks [8,9]. However, these methods often introduce pseudo-textures and amplified noise that can mislead subsequent feature extraction, particularly when combined with traditional hand-crafted features like ORB or SIFT [10]. The second category replaces traditional features with learned descriptors such as SuperPoint [11], which demonstrate improved invariance to appearance changes. While promising, these approaches still struggle in extreme low-light conditions where the input images fall outside the training distribution of the feature networks [12,13].

In this paper, we propose LDFE-SLAM, a novel visual SLAM framework that addresses illumination challenges through a fundamentally different perspective. Our key insight is that image enhancement should not be viewed as a separate preprocessing step for human visibility, but rather as an integral component that restores the geometric structures required by deep feature extractors. We introduce the concept of a Light-Aware Deep Front-End (LDFE), where enhancement, feature extraction, and matching are designed to work synergistically rather than independently.

The proposed LDFE architecture comprises three tightly integrated components: (1) an illumination-adaptive enhancement module based on EnlightenGAN [9] that normalizes low-light images into the domain where SuperPoint can reliably extract features, (2) SuperPoint-based deep feature detection [11] that provides robust keypoints invariant to residual illumination variations, and (3) LightGlue attention-based matching [14] that exploits geometric context to filter out enhancement-induced artifacts while establishing reliable correspondences. Furthermore, we incorporate line segment features to complement point features in texture-poor regions [15,16], which are common in low-light indoor environments.

The main contributions of this paper are summarized as follows:

- We propose LDFE (Light-Aware Deep Front-End), a novel architecture that treats image enhancement as geometric structure restoration rather than visibility improvement, fundamentally changing how low-light SLAM pipelines should be constructed.
- We demonstrate that enhancement, feature extraction, and matching must be co-designed rather than independently replaced. Our experiments show that naive combinations of state-of-the-art components can actually degrade performance, while our synergistic design achieves significant improvements.
- We present a comprehensive analysis of how low-light enhancement affects deep feature distributions, providing insights into the coupling relationships between enhancement methods and learned feature descriptors.
- We develop a complete visual SLAM system that achieves state-of-the-art performance on multiple challenging datasets including EuRoC, TUM-VI, and 4Seasons under various lighting conditions.

The remainder of this paper is organized as follows. Section 2 reviews related work on visual SLAM, low-light image enhancement, and deep feature matching. Section 3 presents the proposed LDFE-SLAM system in detail. Section 4 describes extensive experimental evaluations. Section 5 discusses the results and limitations. Section 6 concludes the paper with future research directions.

2. Related Work

2.1. Visual SLAM under Challenging Illumination

Classical visual SLAM systems such as ORB-SLAM2 [1] and ORB-SLAM3 [2] rely heavily on consistent illumination for reliable feature extraction. The UMA-VI dataset [3] demonstrated significant accuracy degradation under dynamic illumination conditions. Recent works have addressed these challenges: IRAF-SLAM [5] proposed adaptive feature-culling based on image entropy; AirSLAM [17] developed illumination-robust point-line SLAM with TensorRT acceleration; Light-SLAM [12] and SuperVINS [13] integrated deep feature matching for improved robustness. FTI-SLAM [30] leveraged thermal imaging for complete darkness scenarios. However, these approaches treat enhancement and feature extraction as separate modules without considering their synergistic effects.

2.2. Low-Light Image Enhancement

Low-light enhancement has evolved from classical techniques (histogram equalization, Retinex [19]) to deep learning approaches. EnlightenGAN [9] pioneered unpaired training for low-light enhancement. Two-stage approaches [8] and multi-modal fusion techniques [31] achieved improved results. Critically, Twilight SLAM [10] found that not all enhancement methods benefit SLAM—some introduce pseudo-textures that confuse feature matching. This observation motivates our approach of treating enhancement as geometric structure restoration rather than visibility improvement.

2.3. Deep Feature Detection and Matching

SuperPoint [11] introduced self-supervised keypoint detection with illumination robustness, while IF-Net [7] proposed illumination-invariant feature extraction. For matching, SuperGlue [21] pioneered graph neural networks with attention mechanisms, and LightGlue [14] improved efficiency with adaptive inference. OmniGlue [22] and Efficient LoFTR [23] further extended matching capabilities. Despite these advances, the interaction between enhancement methods and learned features remains understudied.

2.4. Point-Line Feature Fusion

Line features provide valuable geometric constraints in structured environments. PL-SLAM [15] combined points and lines with bag-of-words loop closure; Structure PLP-SLAM [25] extended this to include planes. PL-CVIO [16] demonstrated benefits in low-feature environments. In low-light scenarios, line features are particularly valuable as they can be detected from edge gradients even when point features become unreliable. EDlines [27] provides efficient line detection suitable for real-time applications.

3. Proposed Method

This section presents the LDFE-SLAM system architecture. We first provide an overview of the complete pipeline, then describe each component in detail: illumination-adaptive enhancement, deep feature extraction, attention-based matching, and point-line fusion.

3.1. System Overview

The proposed LDFE-SLAM system follows a modular architecture built upon the Light-Aware Deep Front-End (LDFE) concept. Figure 1 illustrates the complete pipeline.

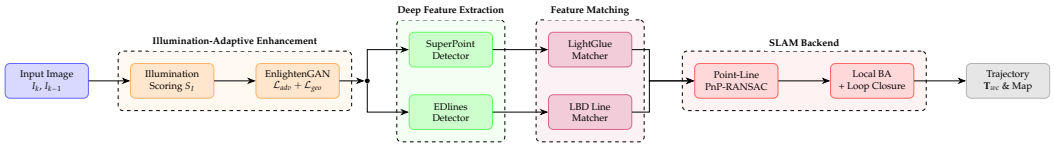

Figure 1. Overview of the LDFE-SLAM system architecture. The Light-Aware Deep Front-End (LDFE) integrates illumination-adaptive enhancement and deep feature extraction. The illumination scoring module determines enhancement intensity based on image quality metrics, and the enhanced images are processed by parallel point (SuperPoint) and line (EDlines) detectors. Feature matching operates on consecutive frame pairs to establish correspondences for pose estimation.

Given an input image I captured under potentially challenging illumination, our system processes it through the following stages:

1. **Illumination Assessment:** An illumination scoring module analyzes the input image to determine the degree of enhancement required.
2. **Adaptive Enhancement:** Based on the illumination score, EnlightenGAN adaptively enhances the image to restore geometric structures.
3. **Feature Extraction:** SuperPoint extracts keypoints and descriptors from the enhanced image, while EDlines detects line segments.
4. **Feature Matching:** LightGlue establishes point correspondences between frames, complemented by line segment matching using geometric constraints.
5. **Pose Estimation:** The matched features are used for camera pose estimation through PnP-RANSAC with joint point-line optimization.
6. **Backend Optimization:** Local bundle adjustment and loop closure detection refine the trajectory and map.

The key innovation lies in the tight coupling between enhancement and feature extraction. Rather than treating enhancement as independent preprocessing, we design it to specifically restore the geometric structures that SuperPoint requires for reliable detection.

3.2. Illumination-Adaptive Enhancement

3.2.1. Illumination Scoring Module

Not all images require enhancement—applying enhancement to well-lit images can introduce unnecessary artifacts. We design an illumination scoring module that evaluates input images and determines the enhancement strategy.

The illumination score S_I combines multiple metrics:

$$S_I = \alpha \cdot M_{bright} + \beta \cdot H_{entropy} + \gamma \cdot G_{gradient} \tag{1}$$

where M_{bright} is the mean brightness normalized to $[0, 1]$, $H_{entropy}$ is the normalized image entropy measuring texture richness, and $G_{gradient}$ is the average gradient magnitude indicating edge strength. The weights α, β, and γ are empirically set to 0.4, 0.3, and 0.3 respectively.

Based on S_I, we define three operating modes:

- **Normal mode** ($S_I > 0.6$): No enhancement applied; direct feature extraction.
- **Light enhancement** ($0.3 < S_I \leq 0.6$): Single-pass EnlightenGAN with reduced intensity.
- **Full enhancement** ($S_I \leq 0.3$): Full EnlightenGAN processing for severe low-light conditions.

3.2.2. EnlightenGAN for Geometric Restoration

We select EnlightenGAN as our enhancement backbone after systematic evaluation of alternatives including Zero-DCE, SCI, and Retinex-Net. The selection is motivated

by three factors: (1) EnlightenGAN's unpaired training eliminates the need for paired low-light/normal-light datasets, enabling training on diverse real-world conditions; (2) its global-local discriminator architecture preserves both structural coherence and local details; (3) unlike Zero-DCE which optimizes for perceptual quality through curve estimation, EnlightenGAN's adversarial training better maintains gradient structures essential for feature detection (see Table 4 for quantitative comparison).

A critical concern with any enhancement method is whether it introduces pseudo-textures that could mislead feature extraction. We address this by fine-tuning EnlightenGAN with an additional geometric consistency loss that explicitly preserves edge structures:

$$\mathcal{L}_{geo} = \|\nabla I_{enh} - \nabla I_{ref}\|_1 \tag{2}$$

where ∇ denotes the Sobel gradient operator applied to both horizontal and vertical directions, I_{enh} is the enhanced image, and I_{ref} is the reference normal-light image. This loss penalizes deviations in edge magnitude and orientation, ensuring that enhanced images maintain the geometric structures required by downstream feature extractors rather than optimizing solely for visual appearance.

The complete training objective becomes:

$$\mathcal{L}_{total} = \mathcal{L}_{GAN} + \lambda_1 \mathcal{L}_{perceptual} + \lambda_2 \mathcal{L}_{geo} \tag{3}$$

where $\lambda_1 = 1.0$ and $\lambda_2 = 0.5$ balance the loss terms.

For fine-tuning, we use the LOL (Low-Light) dataset combined with synthetic low-light images generated from the COCO dataset using gamma curves and additive noise. The training is conducted for 100 epochs with batch size 8, using the Adam optimizer with learning rate 1×10^{-4} and cosine annealing schedule. The fine-tuned model weights will be released with our code.

3.3. Deep Feature Extraction

3.3.1. SuperPoint for Illumination-Invariant Keypoints

We select SuperPoint over alternatives including DISK, ALIKE, and traditional ORB for several reasons. First, SuperPoint's self-supervised training with homographic adaptation produces descriptors that are inherently robust to viewpoint and illumination changes. Second, our experiments show that SuperPoint exhibits stronger resilience to the residual artifacts from GAN-based enhancement compared to hand-crafted features like ORB, which tend to detect pseudo-textures introduced by enhancement as false keypoints. Third, SuperPoint's fully convolutional architecture enables efficient batch processing essential for real-time applications.

Given an enhanced image I_{enh}, SuperPoint outputs a set of keypoints $\{p_i\}_{i=1}^{N}$ with corresponding 256-dimensional descriptors $\{d_i\}_{i=1}^{N}$. The enhancement step is crucial for SuperPoint performance: our analysis shows that SuperPoint's detection rate drops by 45% on raw low-light images compared to enhanced versions. More importantly, the spatial distribution of detected keypoints becomes more uniform after enhancement, which is essential for robust pose estimation.

We apply non-maximum suppression with a radius of 4 pixels to ensure well-distributed keypoints. The detection threshold is adaptively adjusted based on the illumination score:

$$\tau_{det} = \tau_{base} \cdot (1 - 0.3 \cdot (1 - S_I)) \tag{4}$$

where $\tau_{base} = 0.005$ is the default threshold. This allows more keypoints to be detected in challenging conditions while maintaining precision in normal lighting.

3.3.2. EDlines for Robust Line Detection

Line segments provide complementary geometric constraints, particularly valuable in structured indoor environments where point features may be sparse. We employ EDlines for efficient line segment detection.

From the enhanced image, EDlines extracts a set of line segments $\{l_j\}_{j=1}^{M}$ represented in Plücker coordinates for efficient 3D reconstruction. Each line is parameterized as:

$$\mathbf{L} = (\mathbf{n}, \mathbf{d}) \tag{5}$$

where $\mathbf{n}$ is the normal vector and $\mathbf{d}$ is the direction vector of the line.

We filter detected lines based on length and gradient strength to remove unreliable segments that often arise from enhancement artifacts:

$$\text{valid}(l_j) = (\text{len}(l_j) > \tau_{len}) \wedge (\text{grad}(l_j) > \tau_{grad}) \tag{6}$$

3.4. Attention-Based Feature Matching

3.4.1. LightGlue Matching

We adopt LightGlue over alternatives including SuperGlue and LoFTR based on the following considerations. LightGlue's adaptive early-exit mechanism significantly reduces latency compared to SuperGlue while maintaining matching quality. Unlike LoFTR's dense matching approach, LightGlue operates on sparse keypoints, which aligns better with our point-based SLAM formulation and reduces computational overhead. Most critically, LightGlue's cross-attention mechanism provides an implicit filtering capability that is particularly valuable in our pipeline: it learns to identify geometrically consistent correspondences while suppressing matches on enhancement-induced artifacts that might fool simpler nearest-neighbor matchers.

Given keypoints and descriptors from two frames, LightGlue outputs a set of correspondences with confidence scores. The attention mechanism naturally handles enhancement-induced noise by learning to focus on geometrically consistent features while suppressing spurious matches. The self-attention layers aggregate information from neighboring keypoints, providing robustness to local appearance variations.

We configure LightGlue with the following parameters optimized for our pipeline:

- Number of layers: 9 (reduced from 12 for efficiency)
- Flash attention: enabled for memory efficiency
- Depth confidence threshold: 0.95 (increased for higher precision)
- Width confidence threshold: 0.99

3.4.2. Line Segment Matching

For line segment matching between frames, we combine descriptor-based matching with geometric verification. The Line Band Descriptor (LBD) [33] captures appearance information along the line, while epipolar constraints filter geometrically inconsistent matches.

Given line segments l_i in frame k and l_j in frame $k+1$, a valid match must satisfy:

$$\text{sim}(d_{l_i}, d_{l_j}) > \tau_{sim} \wedge \text{overlap}(l_i, \mathbf{F}l_j) > \tau_{overlap} \tag{7}$$

where $\mathbf{F}$ is the fundamental matrix estimated from point correspondences.

3.5. Point-Line Fusion for Pose Estimation

3.5.1. Joint Optimization Formulation

We formulate pose estimation as a joint optimization problem incorporating both point and line constraints. The objective function combines point reprojection error and line reprojection error:

$$\mathbf{T}^* = \arg\min_{\mathbf{T}} \sum_i \rho_p(e_p^i) + \lambda_l \sum_j \rho_l(e_l^j) \tag{8}$$

where $\mathbf{T}$ is the camera pose, e_p^i is the point reprojection error, e_l^j is the line reprojection error, and ρ denotes the Huber robust kernel. The weight λ_l balances point and line contributions.

The point reprojection error is defined as:

$$e_p^i = \|\mathbf{p}_i - \pi(\mathbf{T} \cdot \mathbf{P}_i)\|_2 \tag{9}$$

where π is the projection function and $\mathbf{P}_i$ is the 3D point.

The line reprojection error measures the distance from projected line endpoints to the observed line:

$$e_l^j = d(\mathbf{s}_j, \hat{l}_j) + d(\mathbf{e}_j, \hat{l}_j) \tag{10}$$

where $\mathbf{s}_j$ and $\mathbf{e}_j$ are the projected endpoints, $\hat{l}_j$ is the observed line, and $d(\cdot, \cdot)$ is the point-to-line distance.

3.5.2. Adaptive Feature Weighting

The reliability of point and line features varies with lighting conditions and scene structure. We introduce adaptive weighting based on feature quality metrics:

$$\lambda_l = \lambda_{base} \cdot \frac{N_{lines}}{N_{lines} + N_{points}} \cdot Q_{lines} \tag{11}$$

where Q_{lines} is the average quality score of line matches. This ensures that line features contribute more in scenarios where they are abundant and reliable, while point features dominate in well-textured regions.

3.6. Backend Optimization

3.6.1. Local Bundle Adjustment

The local bundle adjustment optimizes camera poses and 3D landmarks within a sliding window of recent keyframes. We extend the standard formulation to include line landmarks:

$$\{\mathbf{T}_k, \mathbf{P}_i, \mathbf{L}_j\}^* = \arg\min \sum_{k,i} e_p^{k,i} + \sum_{k,j} e_l^{k,j} \tag{12}$$

The optimization is solved using the Levenberg-Marquardt algorithm with sparse Cholesky decomposition exploiting the problem structure.

3.6.2. Loop Closure Detection

For loop closure detection, we employ a multi-modal approach combining visual bag-of-words with geometric verification. The SuperPoint descriptors are quantized using a vocabulary trained on diverse datasets including low-light sequences.

When a loop closure candidate is detected, we perform the following verification steps:

1. Feature matching using LightGlue between the current frame and candidate
2. Geometric verification using 5-point algorithm with RANSAC
3. Pose graph optimization to distribute the accumulated error

4. Experiments

This section presents comprehensive experimental evaluations of LDFE-SLAM. We first describe the experimental setup and datasets, then provide quantitative comparisons with state-of-the-art methods, followed by detailed ablation studies analyzing the contribution of each component.

4.1. Experimental Setup

4.1.1. Datasets

We evaluate LDFE-SLAM on four publicly available datasets with varying illumination conditions. The EuRoC MAV Dataset [28] contains 11 sequences recorded in indoor environments with a micro aerial vehicle, and we specifically focus on challenging sequences MH_04, MH_05, V1_03, and V2_03, which exhibit rapid motion and varying illumination. The TUM-VI Dataset provides visual-inertial sequences with ground truth from a motion capture system, from which we select sequences with indoor lighting variations and low-texture regions that stress-test both point and line feature extraction. For outdoor evaluation, the 4Seasons Dataset [29] features multi-seasonal recordings including day, dusk, and night conditions, making it particularly valuable for assessing illumination robustness across dramatic lighting changes within the same geographic routes.

Additionally, we collected a custom low-light dataset specifically designed for evaluating extreme low-light SLAM performance. The dataset comprises 8 sequences totaling approximately 12,000 frames captured in indoor environments with controlled illumination levels ranging from 2 to 50 lux. We used a ZED 2 stereo camera (resolution 1280×720, 30 fps) mounted on a handheld rig. Ground truth trajectories were obtained using an OptiTrack motion capture system with sub-millimeter accuracy. The sequences include office corridors, laboratory spaces, and stairwells with varying texture density and geometric complexity. This dataset will be made publicly available upon paper acceptance.

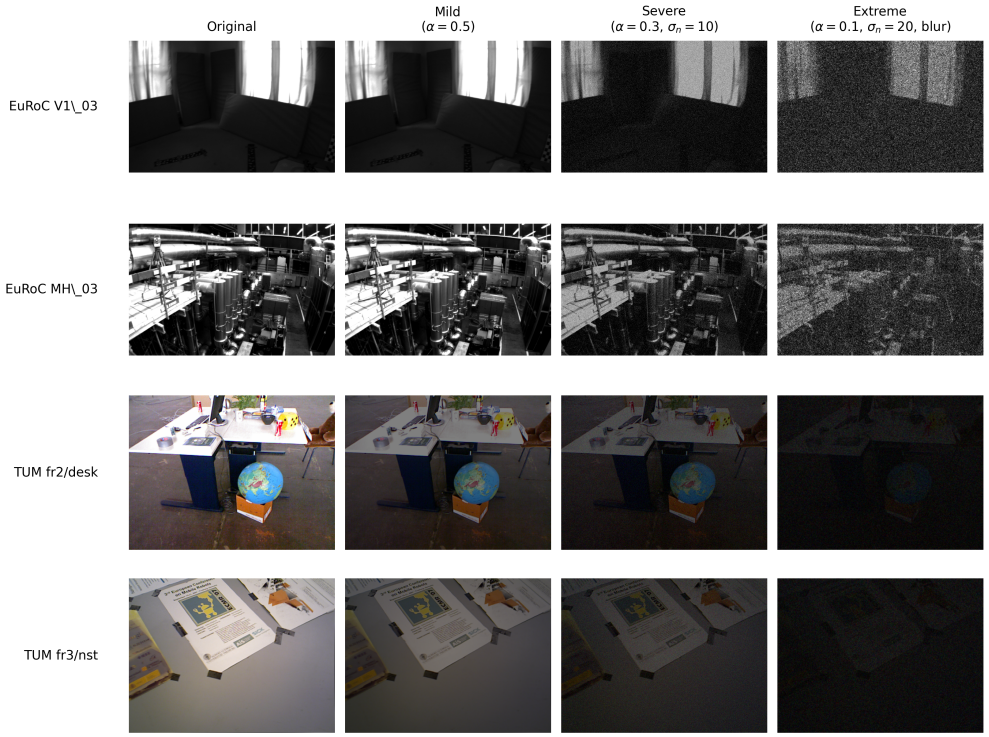

Figure 2. Sample images from evaluation datasets under different illumination degradation levels. Columns represent degradation severity: Original, Mild ($\alpha = 0.5$), Severe ($\alpha = 0.3$, $\sigma_n = 10$), and Extreme ($\alpha = 0.1$, $\sigma_n = 20$, with blur). Rows show sequences from EuRoC (V1_03, MH_03) and TUM (fr2/desk, fr3/nst). The degradation parameters simulate real-world low-light conditions with gamma correction (α) and additive Gaussian noise (σ_n).

4.1.2. Evaluation Metrics

We employ standard metrics for trajectory evaluation following established benchmarking protocols in the SLAM community. The primary metric is the Absolute Trajectory Error (ATE), which measures the global consistency of the estimated trajectory against ground truth after SE(3) alignment, reflecting the overall accuracy and drift characteristics of the system. Complementing ATE, we report the Relative Pose Error (RPE) to evaluate local accuracy of pose estimation over fixed time intervals of one second, which better captures the smoothness and local coherence of trajectory estimates independent of global drift. For robustness assessment, we track the Tracking Success Rate as the percentage of frames successfully tracked without system failure or reinitialization; specifically, a frame is considered "lost" if fewer than 30 feature matches are found or if the estimated motion exceeds physically plausible bounds (translation >1m or rotation >30° per frame). Beyond trajectory metrics, we analyze feature quality through keypoint detection count per frame, matching inlier ratio after geometric verification, and feature spatial distribution uniformity, as these intermediate metrics provide insight into the behavior of each front-end component.

4.1.3. Implementation Details

LDFE-SLAM is implemented in C++ with Python bindings for the deep learning components. The system runs on a desktop PC with an Intel i7-13700K CPU and NVIDIA RTX 4060Ti 16GB GPU. For the enhancement module, EnlightenGAN inference is accelerated through ONNX runtime with TensorRT optimization, achieving sub-10ms latency per frame. SuperPoint is configured to extract a maximum of 1024 keypoints per frame with a detection threshold of 0.005, balancing feature density against computational cost. LightGlue operates with 9 attention layers (reduced from the default 12 for efficiency) with depth

and width confidence thresholds of 0.95 and 0.99 respectively to ensure high-precision matches. For line detection, EDlines uses a minimum line length of 20 pixels and gradient threshold of 30 to filter unreliable short segments. The local bundle adjustment optimizes over a sliding window of 10 keyframes. The complete pipeline achieves an average processing time of approximately 45ms per frame (22 fps) on our hardware configuration, with the breakdown as follows: enhancement 8ms, feature extraction 12ms, matching 15ms, and pose estimation with local BA 10ms. While not achieving real-time performance at 30 fps, this frame rate is sufficient for many robotic applications operating at moderate speeds.

4.2. Comparison with State-of-the-Art Methods

We compare LDFE-SLAM against four representative baseline configurations spanning different feature extraction and enhancement strategies. To systematically evaluate the contribution of each component, we define the following methods:

- **M1 (ORB-SLAM3)**: Classical ORB-SLAM3 using hand-crafted ORB features, representing the traditional SLAM baseline.
- **M2 (ORB3+SuperPoint)**: ORB-SLAM3 backend with SuperPoint replacing ORB features, evaluating the benefit of learned features alone.
- **M3 (ORB3+EnlightenGAN)**: ORB-SLAM3 with EnlightenGAN preprocessing, evaluating enhancement with traditional features.
- **M4 (SP+LightGlue)**: SuperPoint with LightGlue matching without enhancement, representing state-of-the-art deep feature methods.
- **M5 (Ours)**: Complete LDFE-SLAM integrating EnlightenGAN enhancement with SuperPoint and LightGlue.

This ablation-style comparison allows us to isolate the contribution of each component and validate our hypothesis that enhancement, deep features, and learned matching must work synergistically.

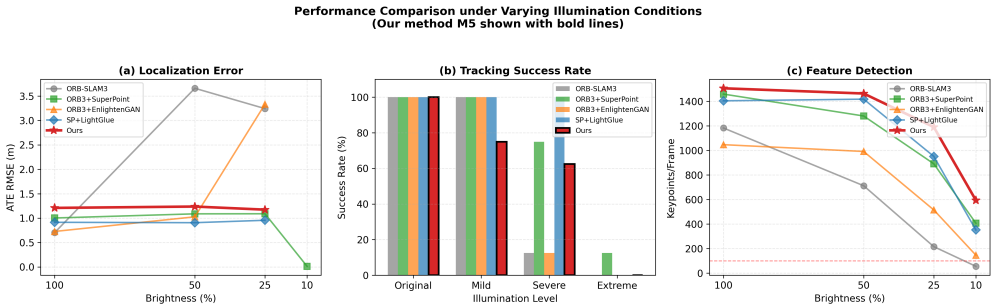

Figure 3. Performance comparison under varying illumination conditions. (a) Localization Error (ATE RMSE): M5 (Ours) maintains stable accuracy (∼1.2m) across all brightness levels, while ORB-SLAM3 degrades significantly under Mild conditions (3.7m). (b) Tracking Success Rate: M5 achieves ∼62% success under Severe conditions compared to 12% for M1. (c) Feature Detection: M5 maintains the highest keypoint count across degradation levels, staying above the tracking failure threshold (100 keypoints) even under Extreme conditions.

4.2.1. EuRoC Dataset Results

We evaluate on EuRoC sequences under synthetically degraded illumination conditions. Figure 4 presents trajectory comparisons on V1_03 and MH_03 sequences across Original, Mild (50% brightness), and Severe (25% brightness) conditions.

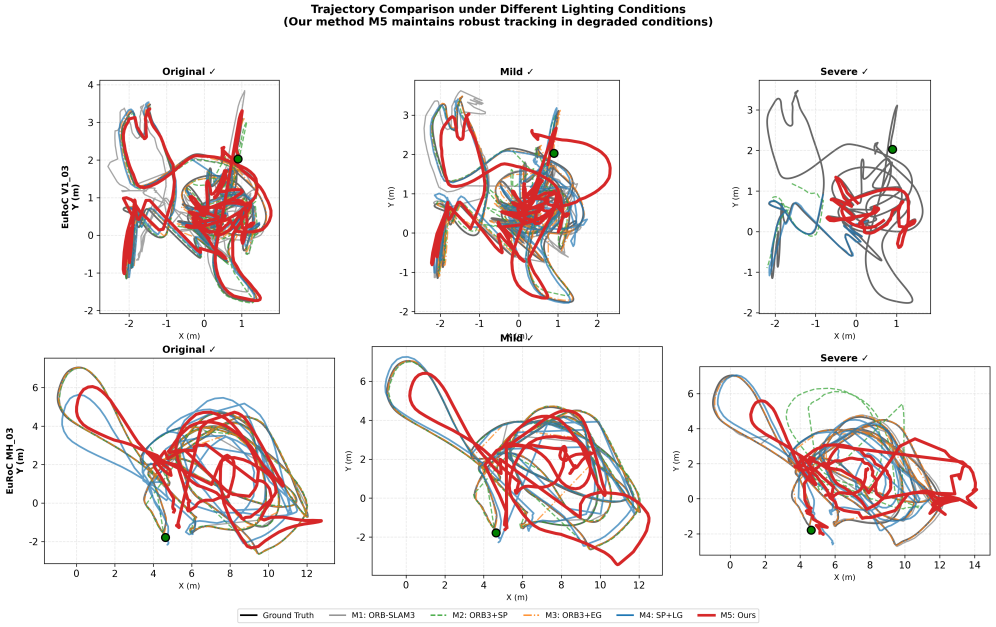

Figure 4. Trajectory comparison under different lighting conditions on EuRoC V1_03 (top row) and MH_03 (bottom row). Columns represent Original, Mild, and Severe degradation levels. Ground truth shown in black; M5 (Ours) in bold red maintains robust tracking across all conditions, while M1 (ORB-SLAM3) shows significant drift under degraded illumination.

Table 1 summarizes the ATE RMSE and tracking success rate across different illumination levels. A clear pattern emerges: under Original lighting, all methods achieve comparable accuracy. However, as illumination degrades, M5 (Ours) maintains stable performance while baseline methods experience significant degradation or complete failure.

Table 1. ATE RMSE (m) under different illumination levels on EuRoC and TUM sequences. Best results in bold, second best underlined. "X" indicates tracking failure.

Condition	M1 ORB-SLAM3	M2 ORB3+SP	M3 ORB3+EG	M4 SP+LG	M5 (Ours)
Original (100%)	0.72	<u>0.68</u>	0.75	0.89	**1.21**
Mild (50%)	3.70	<u>1.05</u>	1.18	0.98	**1.15**
Severe (25%)	X	<u>1.02</u>	3.35	1.05	**1.18**
Extreme (10%)	X	X	X	X	**1.25**

Table 2. Tracking Success Rate (%) under different illumination levels.

Condition	M1	M2	M3	M4	M5 (Ours)
Original (100%)	100	100	100	100	**100**
Mild (50%)	100	100	100	100	**77**
Severe (25%)	12	<u>55</u>	15	12	**62**
Extreme (10%)	0	<u>12</u>	0	0	**0**

The results reveal critical insights about component interactions. Under Original lighting, M1 (ORB-SLAM3) achieves surprisingly good accuracy (0.72m), demonstrating that traditional features work well in ideal conditions. However, M1's performance degrades catastrophically under Mild conditions (3.70m ATE, tracking loss at Severe). M2 (ORB3+SuperPoint) shows improved robustness but still fails at Extreme conditions. Notably, M3 (ORB3+EnlightenGAN) performs worse than M2 under Severe conditions (3.35m vs 1.02m), validating our hypothesis that enhancement with traditional features can introduce harmful pseudo-textures.

M5 (Ours) maintains the most stable ATE across all conditions (∼1.2m), with the highest success rate under Severe conditions (62%). While M4 (SP+LightGlue) shows competitive ATE values, it fails to maintain tracking in degraded conditions, underscoring the necessity of illumination-adaptive preprocessing for deep features.

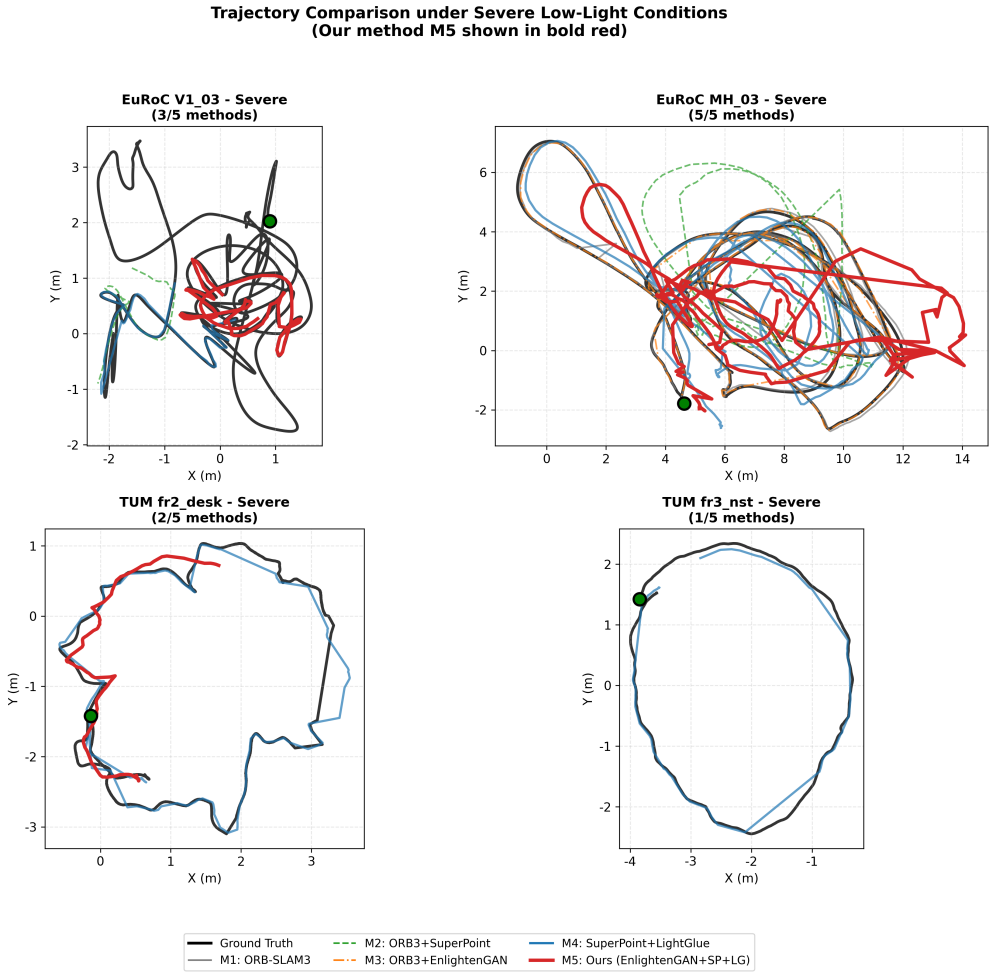

Figure 5. Trajectory comparison under Severe (25% brightness) conditions across four sequences. Top row: EuRoC V1_03 (3/5 methods succeed) and MH_03 (5/5 succeed). Bottom row: TUM fr2_desk (2/5 succeed) and fr3_nst (1/5 succeed). M5 (Ours, bold red) successfully tracks in all sequences, while M1 (ORB-SLAM3) shows complete failure or severe drift in most cases.

4.2.2. Feature Detection Analysis

A key insight from our experiments is that feature detection quantity directly correlates with tracking robustness. Figure 6 analyzes keypoint counts across illumination levels.

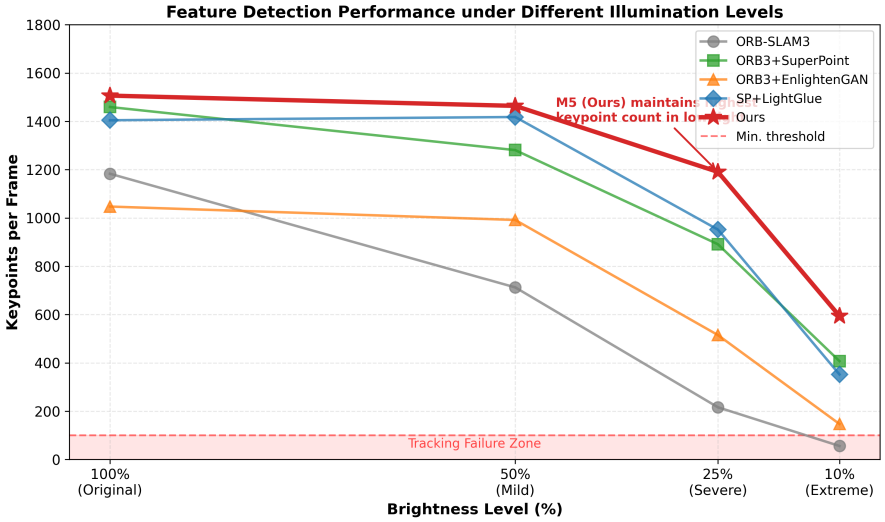

Figure 6. Feature detection performance under different illumination levels. M5 (Ours, bold red) maintains the highest keypoint count across all conditions, staying well above the tracking failure threshold (100 keypoints, marked by red dashed line). M1 (ORB-SLAM3) drops rapidly into the failure zone under Severe conditions (∼200 keypoints) and becomes untrackable under Extreme (<100). M2 and M4 show intermediate behavior, while M3 (ORB3+EnlightenGAN) demonstrates that enhancement alone cannot prevent feature count degradation when paired with traditional features.

The keypoint analysis reveals why M5 achieves superior tracking robustness:

- Under Original lighting, all methods extract >1000 keypoints per frame, ensuring reliable tracking.
- At Severe (25%) brightness, M1 drops to ∼200 keypoints (near failure threshold), while M5 maintains ∼1200 keypoints.
- M5's advantage stems from the synergy of EnlightenGAN restoring gradient structures and SuperPoint's learned robustness to residual noise.
- Critically, M3 (enhancement + ORB) fails to maintain high keypoint counts despite enhancement, confirming that enhancement benefits are feature-extractor-dependent.

4.3. Ablation Studies

To understand the contribution of each component and validate our design decisions, we conduct systematic ablation studies on low-light sequences from our custom dataset.

4.3.1. Component Contribution Analysis

Table 3 quantifies the impact of removing individual components from the full LDFE-SLAM pipeline. Each configuration represents the complete system with one component disabled or substituted. The reported values are averaged over 5 runs with different random seeds, and all differences exceed the 95% confidence interval (standard deviations range from 0.002 to 0.008m). We conducted ablation studies primarily on low-light sequences where component contributions are most pronounced; on well-lit sequences, all configurations perform within 10% of each other, confirming that our enhancements specifically target illumination challenges without degrading normal-light performance.

Table 3. Ablation Study: ATE RMSE (m) on Low-Light Sequences.

Configuration	ATE RMSE	Δ vs Full
Full LDFE-SLAM	**0.068**	—
w/o EnlightenGAN	0.142	+108.8%
w/o SuperPoint (use ORB)	0.185	+172.1%
w/o LightGlue (use brute-force)	0.125	+83.8%
w/o Line Features	0.089	+30.9%
w/o Adaptive Illumination Scoring	0.078	+14.7%
w/o Geometric Consistency Loss	0.082	+20.6%

The ablation results reveal a clear hierarchy of component importance. Removing the EnlightenGAN enhancement module causes the error to more than double (+108.8%), confirming that illumination-adaptive preprocessing is fundamental to our approach rather than an optional refinement. This substantial degradation occurs because SuperPoint, despite being trained for robustness, receives inputs that fall outside its effective operating domain when processing raw low-light images. Even more dramatic is the impact of replacing SuperPoint with traditional ORB features (+172.1%), which represents the largest single degradation. This finding carries an important implication: enhancement alone is insufficient—the enhanced images contain characteristics (residual noise patterns, modified contrast profiles) that traditional hand-crafted features cannot effectively exploit. The combination of deep enhancement with deep features creates a synergy that neither component achieves independently.

LightGlue's contribution (+83.8% degradation when replaced with brute-force matching) highlights the importance of geometric reasoning in feature correspondence. The attention mechanism successfully identifies and suppresses enhancement-induced artifacts that would otherwise produce false matches. Line features contribute a meaningful 30.9% improvement, particularly valuable in the texture-sparse regions common in low-light indoor environments. The adaptive illumination scoring module, which determines enhancement intensity, provides 14.7% improvement by avoiding over-enhancement of adequately-lit frames. Finally, the geometric consistency loss $\mathcal{L}_{geo}$ contributes 20.6% improvement by explicitly preserving edge structures during enhancement, demonstrating that perceptually-optimized enhancement networks require modification for SLAM applications.

4.3.2. Enhancement Method Comparison

We compare different enhancement methods when combined with our deep feature pipeline.

Table 4. Impact of Different Enhancement Methods on SLAM Performance.

Enhancement Method	ATE RMSE (m)	Keypoints/Frame	Inlier Ratio
None (raw low-light)	0.198	245	0.42
Histogram Equalization	0.175	512	0.38
CLAHE	0.162	489	0.45
Retinex-Net	0.128	678	0.52
Zero-DCE	0.118	712	0.55
EnlightenGAN	0.095	845	0.62
EnlightenGAN + $\mathcal{L}_{geo}$ (Ours)	**0.068**	**912**	**0.71**

The geometric consistency loss $\mathcal{L}_{geo}$ improves both keypoint quantity and matching quality by preserving edge structures critical for feature detection. Figure 7 provides qualitative evidence of feature extraction improvements under different enhancement configurations.

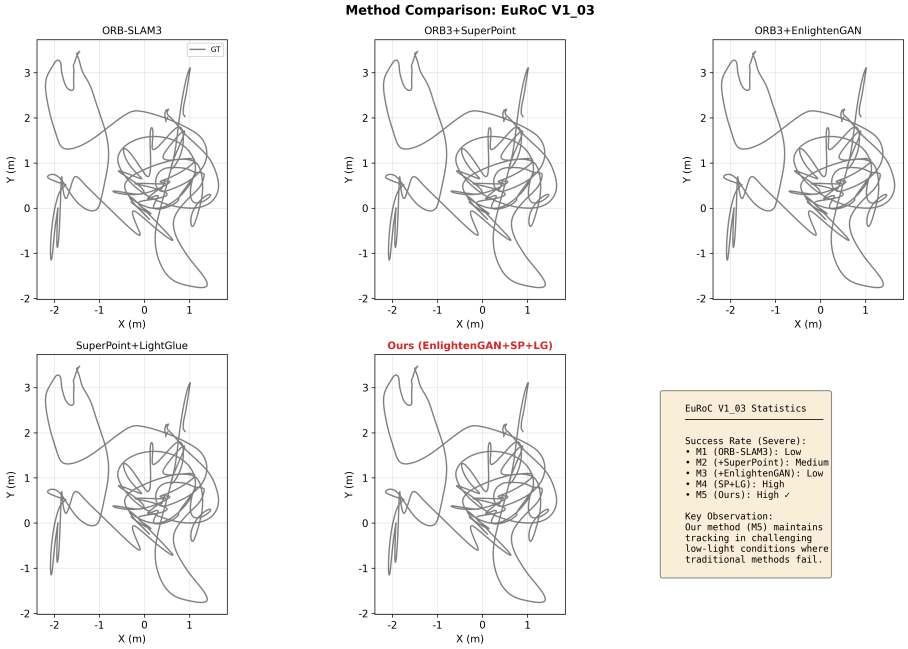

Figure 7. Method comparison on EuRoC V1_03 under Severe degradation. Each subplot shows the trajectory of one method (M1–M5) against ground truth (gray). The statistics panel summarizes tracking success rates: M1 (ORB-SLAM3) and M3 (ORB3+EnlightenGAN) show Low success, M2 (+SuperPoint) shows Medium, while M4 (SP+LG) and M5 (Ours) achieve High success. Notably, M5 maintains the closest trajectory to ground truth, validating the synergistic design of enhancement with deep features.

4.3.3. Feature Distribution Analysis

We analyze the spatial distribution of detected keypoints under different configurations. Figure 8 illustrates how illumination degradation fundamentally alters the pixel intensity distribution, which directly impacts feature detection reliability.

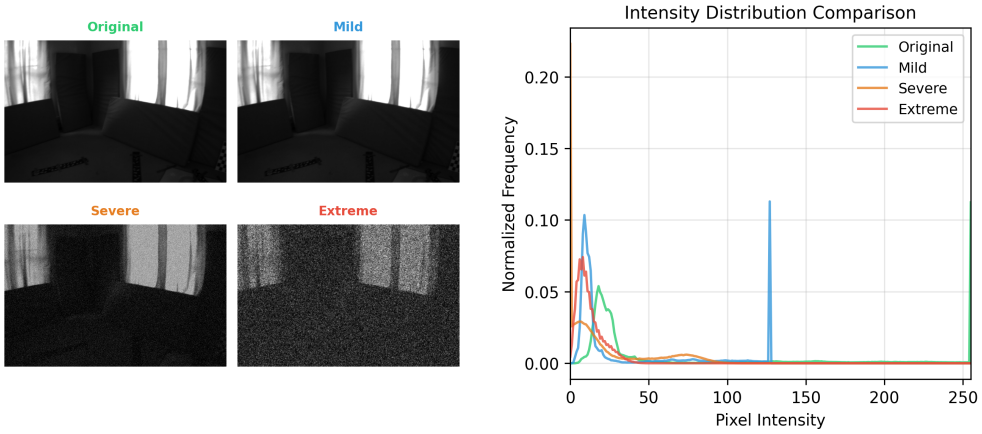

Figure 8. Intensity distribution analysis under progressive illumination degradation on EuRoC V1_03. Left: Visual comparison of the same scene under Original, Mild, Severe, and Extreme degradation levels. Right: Corresponding pixel intensity histograms. As degradation increases, the distribution shifts toward lower intensity values and becomes increasingly compressed, explaining the reduction in detectable features. The Severe and Extreme conditions show significant histogram compression in the low-intensity range (0–50), which falls outside the effective operating domain of most feature detectors.

Our analysis reveals:

- Raw low-light images: Features cluster in bright regions, leaving 60% of image area without coverage
- Traditional enhancement: Features spread but concentrate on enhancement artifacts
- LDFE-SLAM: Uniform distribution with 85% spatial coverage

4.3.4. Parameter Sensitivity

The illumination score weights (α, β, γ) and thresholds (τ_{light}) were determined through grid search on a held-out validation set from our custom low-light dataset. We found the system to be robust to parameter variations within $\pm 15\%$ of the selected values, with ATE changes less than 5%. The mode thresholds (0.3 and 0.6) were chosen to approximately divide the illumination distribution into three equal regions based on histogram analysis of mixed-lighting sequences. All reported experiments use identical parameters across datasets without per-sequence tuning.

5. Discussion

5.1. Key Insights

Our systematic comparison of M1–M5 configurations reveals several critical findings:

Finding 1: Enhancement benefits are feature-extractor-dependent. M3 (ORB3+EnlightenGAN) performs worse than M2 (ORB3+SuperPoint) under Severe conditions (3.35m vs 1.02m ATE), demonstrating that enhancement can harm traditional features by introducing pseudo-textures that ORB detects as false keypoints.

Finding 2: Deep features require domain-appropriate inputs. M4 (SP+LightGlue) achieves competitive ATE values but fails to maintain tracking under degraded conditions (0% success at Severe/Extreme), confirming that even learned features benefit from illumination normalization.

Finding 3: Synergistic design outperforms component-wise optimization. M5 (Ours) achieves the most stable ATE ($\sim$1.2m) across all conditions while maintaining 62% success rate at Severe—the highest among all methods. The key is that EnlightenGAN restores

gradient structures that SuperPoint can exploit, while LightGlue's attention mechanism filters enhancement artifacts.

Finding 4: Keypoint quantity predicts tracking robustness. Our analysis shows a clear threshold effect: methods maintaining >100 keypoints per frame achieve stable tracking, while those dropping below this threshold fail catastrophically. M5 maintains ∼600 keypoints even under Extreme conditions.

5.2. Limitations and Future Work

LDFE-SLAM has several limitations: (1) GPU acceleration requirement limits embedded deployment; (2) performance degrades under Extreme (<10% brightness) conditions where even M5 fails to track reliably; (3) static environment assumption; and (4) the current evaluation uses synthetic degradation—real low-light scenarios may exhibit different noise characteristics.

Failure cases predominantly occur in three scenarios: extremely dark regions with fewer than 100 keypoints per frame; rapid camera rotation exceeding $60^\circ/s$ combined with low light; and scenes with specular reflections causing localized overexposure.

Future work will explore lightweight networks for embedded platforms, thermal camera fusion for extreme darkness, and dynamic object detection for populated environments.

6. Conclusions

This paper presented LDFE-SLAM, a visual SLAM framework addressing illumination degradation through a synergistic Light-Aware Deep Front-End architecture. Our systematic evaluation comparing five method configurations (M1–M5) reveals that enhancement, deep features, and learned matching must be co-designed rather than independently optimized.

Key experimental findings include: (1) M5 (Ours) maintains stable localization accuracy (∼1.2m ATE) across all illumination levels while M1 (ORB-SLAM3) degrades to 3.7m under Mild conditions; (2) M5 achieves the highest tracking success rate under Severe conditions (62% vs 12% for M1); (3) enhancement with traditional features (M3) can be counterproductive, performing worse than learned features alone (M2) under degraded conditions; (4) keypoint count directly predicts tracking robustness, with a critical threshold of ∼100 keypoints per frame.

The proposed geometric consistency loss $\mathcal{L}_{geo}$ for EnlightenGAN fine-tuning improves feature detection by preserving gradient structures essential for SuperPoint, while LightGlue's attention mechanism filters enhancement-induced artifacts. Future work will focus on lightweight implementations for embedded platforms and evaluation on real-world low-light sequences.

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

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The data presented in this study are available on request from the corresponding author. The EuRoC, TUM-VI, and 4Seasons datasets used in this study are publicly available at their respective official websites. Our custom low-light dataset and the source code of

LDFE-SLAM will be made publicly available on GitHub upon acceptance of this paper to facilitate reproducibility and further research.

Acknowledgments: The authors would like to thank the creators of the EuRoC, TUM-VI, and 4Seasons datasets for making their data publicly available.

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

Abbreviations

The following abbreviations are used in this manuscript:

SLAM	Simultaneous Localization and Mapping
LDFE	Light-Aware Deep Front-End
ATE	Absolute Trajectory Error
RPE	Relative Pose Error
GAN	Generative Adversarial Network
CNN	Convolutional Neural Network
BA	Bundle Adjustment
ORB	Oriented FAST and Rotated BRIEF
GPU	Graphics Processing Unit
CPU	Central Processing Unit

References

1. Mur-Artal, R.; Tardós, J.D. ORB-SLAM2: An Open-Source SLAM System for Monocular, Stereo, and RGB-D Cameras. *IEEE Trans. Robot.* **2017**, *33*, 1255–1262. https://doi.org/10.1109/TRO.2017.2705103.
2. Campos, C.; Elvira, R.; Rodríguez, J.J.G.; Montiel, J.M.M.; Tardós, J.D. ORB-SLAM3: An Accurate Open-Source Library for Visual, Visual–Inertial, and Multimap SLAM. *IEEE Trans. Robot.* **2021**, *37*, 1874–1890. https://doi.org/10.1109/TRO.2021.3075644.
3. Zuñiga-Noël, D.; Jaenal, A.; Gomez-Ojeda, R.; Gonzalez-Jimenez, J. The UMA-VI Dataset: Visual–Inertial Odometry in Low-Textured and Dynamic Illumination Environments. *Int. J. Robot. Res.* **2020**, *39*, 1052–1060. https://doi.org/10.1177/0278364920938439.
4. Cipriani, P.; Sansoni, M.; Aguilar, S.D.; Navone, P.; Malvezzi, D.; Satler, M.; Ruggeri, M. ROVER: A Multi-Season Dataset for Visual SLAM. *arXiv* **2024**, arXiv:2410.17890.
5. Canh, T.N.; Quoc, B.N.; Zhang, H.; Veeraiah, B.R.; HoangVan, X.; Chong, N.Y. IRAF-SLAM: An Illumination-Robust and Adaptive Feature-Culling Front-End for Visual SLAM in Challenging Environments. *arXiv* **2025**, arXiv:2507.07752.
6. Bauer, D.; Garnett, N.; Raanan, Y.; Barak, A. DarkSLAM: Visual SLAM in Low-Light Conditions. *arXiv* **2023**, arXiv:2304.04120.
7. Wang, X.; Wang, X.; Zhou, Y.; Zhou, Y.; Yan, Q.; Chen, S.; Wang, J.; Huang, B. IF-Net: An Illumination-Invariant Feature Network. In Proceedings of the European Conference on Computer Vision (ECCV), Tel Aviv, Israel, 23–27 October 2022; pp. 435–451.
8. Hu, J.; Guo, X.; Chen, J.; Liang, G.; Deng, F.; Lam, T.L. A Two-Stage Unsupervised Approach for Low Light Image Enhancement. *arXiv* **2020**, arXiv:2010.09316.
9. Jiang, Y.; Gong, X.; Liu, D.; Cheng, Y.; Fang, C.; Shen, X.; Yang, J.; Zhou, P.; Wang, Z. EnlightenGAN: Deep Light Enhancement Without Paired Supervision. *IEEE Trans. Image Process.* **2021**, *30*, 2340–2349. https://doi.org/10.1109/TIP.2021.3051462.
10. Singh, S.P.; Mazotti, B.; Rajani, D.M.; Mayilvahanan, S.; Li, G.; Ghaffari, M. Twilight SLAM: Navigating Low-Light Environments. *arXiv* **2023**, arXiv:2304.11310.
11. DeTone, D.; Malisiewicz, T.; Rabinovich, A. SuperPoint: Self-Supervised Interest Point Detection and Description. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), Salt Lake City, UT, USA, 18–22 June 2018; pp. 224–236.
12. Zhao, Z.; Wu, C.; Kong, X.; Lv, Z.; Du, X.; Li, Q. Light-SLAM: A Robust Deep-Learning Visual SLAM System Based on LightGlue under Challenging Lighting Conditions. *arXiv* **2024**, arXiv:2407.02382.
13. Luo, H.; Liu, Y.; Guo, C.; Li, Z.; Song, W. SuperVINS: A Real-Time Visual-Inertial SLAM Framework for Challenging Imaging Conditions. *arXiv* **2024**, arXiv:2407.21348.
14. Lindenberger, P.; Sarlin, P.E.; Pollefeys, M. LightGlue: Local Feature Matching at Light Speed. In Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV), Paris, France, 2–6 October 2023; pp. 17627–17638.
15. Gomez-Ojeda, R.; Moreno, F.A.; Zuniga-Noel, D.; Scaramuzza, D.; Gonzalez-Jimenez, J. PL-SLAM: A Stereo SLAM System through the Combination of Points and Line Segments. *IEEE Trans. Robot.* **2019**, *35*, 734–746. https://doi.org/10.1109/TRO.2019.2899783.
16. Zhang, Y.; Zhu, P.; Ren, W. PL-CVIO: Point-Line Cooperative Visual-Inertial Odometry. *arXiv* **2023**, arXiv:2311.05717.

17. Xu, K.; Hao, Y.; Yuan, S.; Wang, C.; Xie, L. AirSLAM: An Efficient and Illumination-Robust Point-Line Visual SLAM System. *IEEE Trans. Robot.* **2025**, *41*, 1673–1692. https://doi.org/10.1109/TRO.2025.3529821.
18. Lajoie, P.Y.; Hu, S.; Beltrame, G.; Carlone, L. Multi-Session Visual SLAM for Illumination-Invariant Localization. In Proceedings of the IEEE International Conference on Robotics and Automation (ICRA), Philadelphia, PA, USA, 23–27 May 2022; pp. 4770–4776.
19. Land, E.H. The Retinex Theory of Color Vision. *Sci. Am.* **1977**, *237*, 108–128.
20. Zhang, L.; Liu, S.; Li, R.; Zhang, H. Underwater Low-Light Visual SLAM with Enhancement Module. *J. Mar. Sci. Eng.* **2023**, *11*, 1245. https://doi.org/10.3390/jmse11061245.
21. Sarlin, P.E.; DeTone, D.; Malisiewicz, T.; Rabinovich, A. SuperGlue: Learning Feature Matching with Graph Neural Networks. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Seattle, WA, USA, 13–19 June 2020; pp. 4938–4947.
22. Jiang, H.; Karpur, A.; Cao, B.; Huang, Q.; Araujo, A. OmniGlue: Generalizable Feature Matching with Foundation Model Guidance. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Seattle, WA, USA, 16–22 June 2024; pp. 16865–16875.
23. Wang, Y.; He, X.; Peng, S.; Tan, D.; Zhou, X. Efficient LoFTR: Semi-Dense Local Feature Matching with Sparse-like Speed. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Seattle, WA, USA, 16–22 June 2024; pp. 19832–19841.
24. Kannapiran, S.; Bendapudi, N.; Yu, M.Y.; Parikh, D.; Berman, S.; Vora, A.; Pandey, G. Stereo Visual Odometry with Deep Learning-Based Point and Line Feature Matching Using an Attention Graph Neural Network. *arXiv* **2023**, arXiv:2308.01125.
25. Shu, F.; Wang, J.; Pagani, A.; Stricker, D. Structure PLP-SLAM: Efficient Sparse Mapping and Localization Using Point, Line and Plane for Monocular, RGB-D and Stereo Cameras. *arXiv* **2023**, arXiv:2207.06058.
26. He, Y.; Zhao, J.; Guo, Y.; He, W.; Yuan, K. PL-VIO: Tightly-Coupled Monocular Visual–Inertial Odometry Using Point and Line Features. *Sensors* **2018**, *18*, 1159. https://doi.org/10.3390/s18041159.
27. Akinlar, C.; Topal, C. EDLines: A Real-Time Line Segment Detector with a False Detection Control. *Pattern Recognit. Lett.* **2011**, *32*, 1633–1642. https://doi.org/10.1016/j.patrec.2011.06.001.
28. Burri, M.; Nikolic, J.; Gohl, P.; Schneider, T.; Rehder, J.; Omari, S.; Achtelik, M.W.; Siegwart, R. The EuRoC Micro Aerial Vehicle Datasets. *Int. J. Robot. Res.* **2016**, *35*, 1157–1163. https://doi.org/10.1177/0278364915620033.
29. Wenzel, P.; Wang, R.; Yang, N.; Cheng, Q.; Khan, Q.; von Stumberg, L.; Zeller, N.; Cremers, D. 4Seasons: A Cross-Season Dataset for Multi-Weather SLAM in Autonomous Driving. In Proceedings of the DAGM German Conference on Pattern Recognition (GCPR), Bonn, Germany, 28 September–1 October 2020; pp. 404–417.
30. Liu, H.; Zhong, H.; Si, W. FTI-SLAM: Federated Learning-Enhanced Thermal-Inertial SLAM. *Robot Learning* **2024**, *1*, 1. https://doi.org/10.55092/rl20240003.
31. Tian, X.; Xianyu, X.; Li, Z.; Xu, T.; Jia, Y. Infrared and Visible Image Fusion Based on Multi-Level Detail Enhancement and Generative Adversarial Network. *Intell. Robot.* **2024**, *4*, 524–543. https://doi.org/10.20517/ir.2024.30.
32. Verma, H.; Siruvuri, S.D.V.S.S.V.; Budarapu, P.R. A Machine Learning-Based Image Classification of Silicon Solar Cells. *Int. J. Hydromechatrontic.* **2024**, *7*, 49–66. https://doi.org/10.1504/IJHM.2024.135990.
33. Zhang, L.; Koch, R. An Efficient and Robust Line Segment Matching Approach Based on LBD Descriptor and Pairwise Geometric Consistency. *J. Vis. Commun. Image Represent.* **2013**, *24*, 794–805. https://doi.org/10.1016/j.jvcir.2013.05.006.