remote sensing

MDPI

Article

Consistent Photometric Enhancement Network for Remote Sensing Change Detection

Pengcheng Han [1,2,3], Zhenyu Xia [1,2], Lin Chen [1,2], Boni Hu [1,2], Naoufel Werghi [3] and Shuhui Bu [1,2,*]

[1] College of Aeronautics, Northwestern Polytechnical University, Xi'an 710072, China; hanpc1125@nwpu.edu.cn (P.H.); xiazhenyu@mail.nwpu.edu.cn (Z.X.); npuchenlin@mail.nwpu.edu.cn (L.C.); huboni@mail.nwpu.edu.cn (B.H.)
[2] National Key Laboratory of Aircraft Configuration Design, Xi'an 710072, China
[3] Department of Computer Science, Khalifa University, Abu Dhabi P.O. Box 127788, United Arab Emirates; naoufel.werghi@ku.ac.ae
* Correspondence: bushuhui@nwpu.edu.cn

Highlights

What are the main findings?

- Remote sensing change detection involves multi-temporal data acquired across large temporal and spatial ranges, where illumination, weather, and cloud conditions often cause uneven and inconsistent photometric distributions.
- CPEN reduces these photometric discrepancies through low-light compensation and joint bi-temporal enhancement, achieving the best F1 and IoU on LEVIR-CD, PRCV-CD, SYSU-CD, and NPULL-CD compared with recent baselines.

What are the implications of the main findings?

- Independently enhancing images with different photometric distributions may further increase cross-temporal inconsistency, indicating that task-oriented photometric consistency is more effective for low-light change detection than separate image enhancement.
- CPEN improves the reliability of remote sensing change detection in practical scenarios with low illumination, inconsistent brightness, and complex lighting conditions.

Abstract

Change detection performance in remote sensing is highly sensitive to illumination variations between bi-temporal images. In real-world scenarios, low-light conditions and inconsistent brightness often lead to degraded feature representations and unreliable detection results. To address this issue, this paper proposes a Consistent Photometric Enhancement Network (CPEN) for change detection under complex illumination conditions. Unlike conventional low-light enhancement methods applied independently to each image, CPEN explicitly enforces photometric consistency between bi-temporal images before and during an enhancement procedure. The proposed framework applies a low-light compensation mechanism to reduce photometric discrepancies between image pairs, which is followed by a bi-temporal enhancement module that jointly improves brightness while preserving shared structural information. Using the photometrically consistent and enhanced images, a change detection module is employed to extract reliable features and generate accurate change maps. Experimental results show that CPEN achieves F1/IoU scores of 90.39/82.67, 80.71/70.58, 78.34/64.26, and 86.26/79.02 on LEVIR-CD, PRCV-CD, SYSU-CD, and the real-world NPULL-CD dataset, respectively, demonstrating its robustness under both simulated and real illumination variations.

Check for updates

Academic Editor: Zhenming Peng

Received: 9 June 2026
Revised: 17 July 2026
Accepted: 21 July 2026
Published: 4 August 2026

Copyright: © 2026 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license.

Keywords: change detection; photometric enhancement; multi-brightness; complex illumination conditions; remote sensing

1. Introduction

Change detection identifies differences in the same area at different times using a pair of remote sensing images. With the rapid development of hardware capabilities and algorithm technology, it is easy to acquire and analyze large amounts of remote sensing images. Therefore, the change detection task has become a promising and fundamental task in the remote sensing community, and plays a key role in the fields of urbanization monitoring [1], land use change detection [2], disaster assessment [3], agricultural surveys and ecosystem monitoring [4].

Existing remote sensing change detection methods can be discussed from several perspectives rather than a single traditional taxonomy. From the analysis granularity perspective, they include scene-level, object-level, and pixel-level methods [5]. From the processing framework perspective, representative methods include image differencing/transformation methods, post-classification comparison, and recent deep learning-based methods such as CNN-, Transformer-, and Mamba-based frameworks [6–9]. In particular, post-classification comparison detects changes by comparing independently classified bi-temporal maps, and its reliability depends strongly on the classification accuracy of each temporal image rather than directly extracting pairwise change features. In contrast, recent deep models usually learn feature interaction or feature alignment from paired bi-temporal inputs. Despite these differences, most change detection methods implicitly assume that the two temporal images are sufficiently comparable in radiometric and illumination conditions.

With the widespread use of satellites and drones, remote sensing images are increasingly acquired under complex illumination conditions, including uneven light, weak illumination, and backlight. These conditions violate the radiometric comparability assumption of many change detection methods and motivate the following discussion of low-light image enhancement. In this paper, low-light enhancement is therefore introduced not as a standalone image restoration objective, but as a way to improve bi-temporal photometric comparability for downstream change detection.

Low-light image enhancement technology, which aims to recover the desired light status in degraded regions, has attracted wide attention recently. These innovative algorithms proposed to address the ill-posed enhancement problem could be categorized into two groups: (1) conventional model-based methods, which enhance the degradation regions according to a physical model but require massive expert prior knowledge; (2) deep learning-based methods, which learn mapping rules from dark to bright based on large amounts of data. However, the performance of directly applying these techniques to change detection task is not ideal. The core problem is that when these brightness enhancement methods are applied to images with different brightness, there is an obvious inconsistency in the brightness distribution of the enhanced images. But the common assumption is that the light distribution between compared sequential images or across training and test datasets is similar for change detection tasks.

As shown in Figure 1, the brightness distribution of images acquired at two different times has obvious differences. After enhancement of the current general enhancement network [10–14], the enhanced images also show large differences in luminosity distribution, which will bring many uncertainties to downstream change detection tasks.

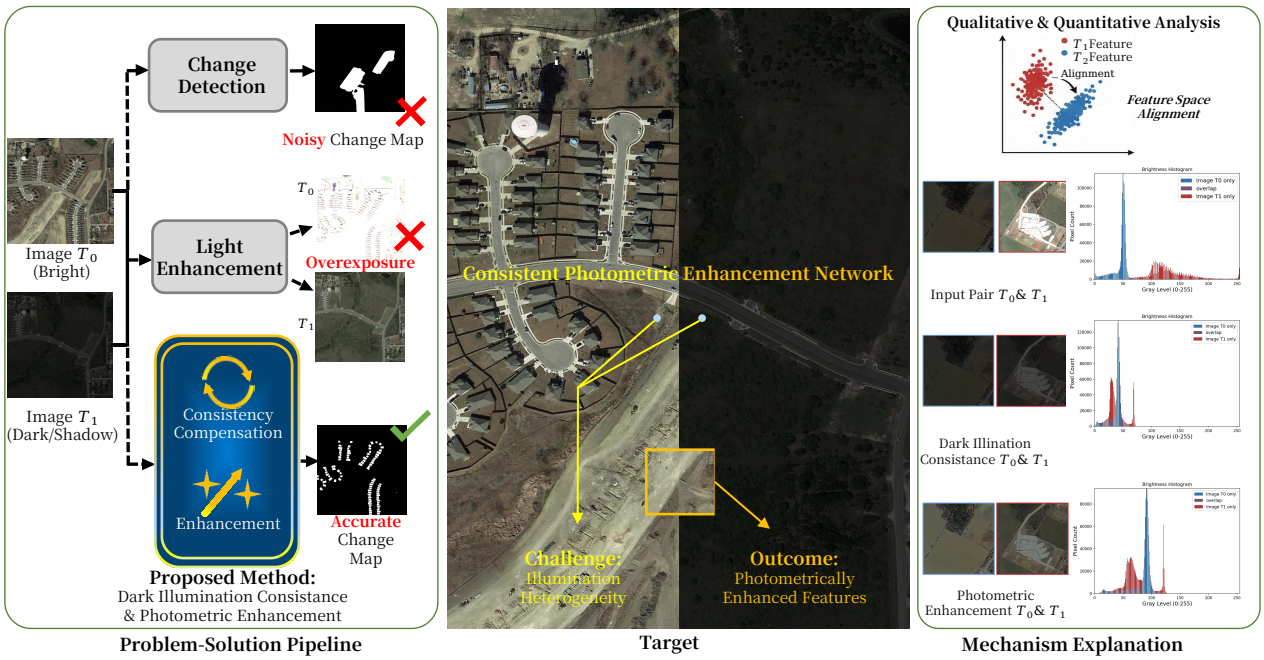


Figure 1. In the presence of photometric inconsistency, directly applying conventional difference detection algorithms or generic photometric enhancement methods often leads to insufficient texture feature learning in low-light regions and overexposure artifacts, thereby degrading the detection performance. To address these challenges, this work achieves feature-space alignment between image pairs through dark-illumination compensation and photometric enhancement, which leads to accurate difference detection under complex and varying illumination conditions.

It is important to emphasize that the target problem addressed in this paper is not generic low-light image enhancement, but bi-temporal change detection under inconsistent illumination. Different from conventional "enhancement + detection" pipelines that process each image independently and then feed the restored images to a detector, our method explicitly constrains cross-temporal photometric consistency before and during enhancement. This pairwise consistency design is the key distinction of CPEN because, for change detection, the absolute visual quality of each image alone is insufficient; what matters more is whether the two enhanced images remain comparable in brightness distribution, structural response, and downstream feature space.

The impact of inconsistent brightness on change detection feature extraction is mainly reflected in the following two aspects: (1) The negative impact of inconsistent brightness changes on compared bi-temporal image quality: Inconsistent brightness changes will cause changes in the contrast of the image, which will affect the overall quality of the image. For example, an image that is too bright or too dark will cause details to be lost, making feature extraction difficult. (2) The negative impact on the accuracy of feature matching: Inconsistent brightness changes will cause the pixel values of the images to be unevenly distributed, resulting in the feature extraction and matching algorithm being unable to accurately extract stable features for compared bi-temporal images.

Therefore, to prevent a change detection task from being negatively affected by complex luminosity, such as inconsistent brightness and low brightness, we propose a consistent photometric enhancement network (CPEN) for change detection. As shown in [15], neural networks tend to learn low-frequency signals first, and some high-frequency components may be discarded during the reconstruction and rendering process. The framework utilizes these neural network characteristics to design a dark-light compensation mechanism to achieve dark-light consistency for bi-temporal images. Then, a brightness enhancement module is designed to achieve consistent brightness enhancement, which will further facili-

tate the extraction of difference detail features by the change detection network. In this way, the framework could effectively solve the problem of a change detection network failing in low brightness conditions or producing poor results in inconsistent brightness conditions. Compared with existing enhancement-assisted change detection pipelines, the novelty of CPEN lies in three aspects: (1) a low-light compensation mechanism is introduced to explicitly reduce cross-temporal photometric discrepancy before enhancement; (2) a bi-temporal enhancement module jointly restores image brightness while preserving shared structures across the compared image pair; (3) the whole framework is optimized for downstream change detection rather than standalone image restoration quality. Our main contributions are summarized as follows.

- We study the task of change detection under complex illumination conditions and analyze the problem of change detection failure due to inconsistent brightness and low brightness.
- The dark-light compensation mechanism module is designed to reconstruct consistent dark light for bi-temporal images by fully learning low-frequency signals and suppressing high-frequency signals.
- The end-to-end principled brightness enhancement network is proposed for change detection. The constraints between two compared images could preserve shared features across different brightness domains, and the enhanced module could effectively increase the brightness of the compared images.

To validate the effectiveness of the proposed method, a multi-brightness change detection dataset is designed and created, which includes three geographically distinct environments with varying acquisition brightness conditions. The method demonstrates superior performance compared to existing low-light enhancement methods and state-of-the-art CD methods, as supported by the results of extensive experiments.

2. Related Work

This research is focused on two related areas: change detection and low-light enhancement. This part involves the discussion and analysis of the latest research findings in both fields.

2.1. Change Detection

Remote sensing change detection technology is used to discover and identify differences in ground target areas with two or more consecutive sequences of images from the same geographic location at different times. From the perspective of the scale level of analysis, change detection methods are mainly divided into three aspects [5]: scene-level methods, object-level and pixel-level methods.

Scene-level change detection in remote sensing refers to the process of recognizing land use changes by analyzing multi-temporal remote sensing images of a specific area. This approach leverages feature information from the input images to create a change map and uses classification methods to identify the presence of changes within distinct patches [16–18]. The object-level approach [19–21] prioritizes the identification of the "changed area" as the primary focus of detection in remote sensing images, distinguishing it from the "unchanged area," which serves as the background. This strategy is commonly applied in change detection for high-resolution remote sensing images. Additionally, standard object detection techniques [19] can be effectively utilized for this purpose. Generally, pixel-based change detection methods analyze features from individual pixels along with their surrounding context. These techniques work by generating binary masks that classify each pixel as either changed or unchanged [6,22,23].

Recent change detection studies have shifted from hand-crafted differencing and simple feature fusion to stronger bi-temporal representation learning. Early FCN- and Siamese-based networks explored how to fuse features from paired remote sensing images [6,24], while recent CNN-based models further strengthen multilevel temporal interaction and spatiotemporal feature fusion [25,26]. Transformer-based methods introduce long-range contextual modeling and token-level temporal interaction for remote sensing change detection [7,8,27]. More recently, Mamba-based models have been introduced to combine local clues with efficient global dependency modeling [9], and incremental methods attempt to adapt change detection models to evolving data domains [28]. Recent advances further broaden the problem setting by addressing arbitrary resolution differences with spatial-frequency dual-domain learning [29] and formulating multimodal change detection from an image-matching perspective [30]. These advances improve semantic discrimination and temporal interaction, but many supervised detectors for homogeneous optical images still assume that the bi-temporal inputs are radiometrically comparable. When low-light degradation or brightness inconsistency changes the feature response of unchanged regions, the detector may learn illumination-induced differences as false changes. This limitation motivates illumination-aware pairwise photometric normalization before change discrimination.

Remote sensing change detection methods are rapidly evolving, with current research aimed at improving detection accuracy for ideal scenes. However, real-world scenes often feature data with varying complex lighting, which can lead to failures in existing approaches under challenging brightness conditions. And low-light image enhancement (LLIE) techniques may help mitigate this issue to some degree.

2.2. Low-Light Image Enhancement

Image enhancement is a key task in image processing, which is designed to amplify the differences between features of various objects in images, minimize irrelevant features, improve image quality, increase the amount of information, enhance image interpretation and recognition, and fulfill specific analytical needs. Existing enhancement methods are mainly divided into traditional methods and deep learning-based LLIE.

Traditional methods for low-light enhancement include Histogram Equalization (HE) [31], Gamma Correction (GC) [32] methods and Retinex model-based [33,34] methods. The HE and GC methods directly enhance the low visibility and contrast of underexposed images while minimally accounting for illumination factors, resulting in enhanced images that may appear perceptually inconsistent with real scenes illuminated by normal light. They are based on the Retinex theory, treating the reflectance component of low-light images as a viable solution for enhancement. For instance, Guo et al. [35] suggest refining the initially estimated illumination map by applying a structural prior. However, these methods have several limitations: (1) they assume low-light images are free from corruption, which can result in significant noise and color distortion during enhancement; (2) they depend on hand-crafted priors that often require meticulous parameter adjustments and exhibit limited generalization; (3) their complex optimization process leads to relatively long runtime.

In recent years, deep learning-based approaches to low-light image enhancement (LLIE) have achieved remarkable success since the initial groundbreaking work [10–13]. These solutions offer improved accuracy, robustness, and speed compared to traditional methods, garnering increasing interest. For example, Wei et al. [36] and subsequent studies [37] integrate Retinex decomposition with deep learning techniques. However, these approaches often involve a complex multi-stage training process, utilizing several convolutional neural networks (CNNs) to learn or adjust various components of the Retinex

model separately. Wang et al. [38] introduced a one-stage Retinex-based CNN called DeepUPE, which aims to directly predict the illumination map. Despite this, DeepUPE overlooks corruption factors, resulting in heightened noise and color distortion when enhancing underexposed images. Additionally, CNN-based methods tend to struggle with capturing long-range dependencies across different regions.

Thus, we need to consider how to effectively extract both the common feature representations across varying brightness levels and the unique characteristics of specific domains in change detection to mitigate the adverse effects of complex lighting conditions. Cai et al. [14] propose the first transformer-based algorithm RetinexFormer for low-light image enhancement, which designs a new self-attention mechanism to utilize the illumination information as a key clue to guide the modeling of long-range dependencies. These methods work well in feature extraction and long-term modeling to make the enhancement method more natural. From the perspective of illumination-robust change detection, structure-consistency-based methods provide another important strategy for mitigating radiometric, illumination, and modality-induced interference. For example, ALSC constructs adaptive local graphs to exploit local structure consistency between heterogeneous images [39]. IRG-McS further introduces an iterative robust graph framework for unsupervised heterogeneous change detection [40]. More recently, RIEM formulates multimodal change detection as a rules-induced energy model and detects changes by analyzing intra-image relationships rather than relying on direct inter-image radiometric comparison [41]. These methods are robust because structural relationships are usually more stable than raw pixel intensities under different imaging conditions. However, they mainly focus on structure-level invariance for heterogeneous or multimodal change detection. In contrast, CPEN addresses homogeneous optical bi-temporal images under low-light and brightness-inconsistent conditions by explicitly learning pairwise photometric compensation and enhancement before change discrimination. Therefore, CPEN is complementary to structure-consistency-based methods: the former reduces illumination-induced photometric discrepancy, whereas the latter emphasizes modality-invariant structural relationships.

Moreover, most existing low-light enhancement approaches are designed for single-image visual restoration and do not explicitly constrain the consistency of bi-temporal image pairs. As a result, applying them independently to two remote sensing images may improve visibility but still enlarge cross-temporal photometric discrepancy, which is harmful to subsequent change detection. In contrast, our work in CPEN focuses on learning a consistent photometric distribution with light compensation and photometric constraints between two compared images for change detection tasks. Therefore, CPEN should be understood as a task-oriented pairwise photometric normalization and enhancement framework, rather than a simple serial combination of enhancement and detection.

3. Materials and Methods

3.1. Problem Formulation

Change detection. Given a pair of images $I_{db} = \{I_{db}^1, I_{db}^2\}$ from different brightness $\mathcal{D}_b$, the objective of change detection is to predict a binary map with 1 and 0 to indicate change and no change. Generally, brightness will change over time. For binary pixel classification, cross-entropy is a standard supervised objective in deep networks [42]. Thus, a single-step detection model is commonly optimized by a pixel-wise cross-entropy loss. The network can be decomposed into a feature extractor and a classifier. The former generates a dense feature map $F \in \mathbb{R}^{H\times W\times d}$, where H, W and d denote the height, width and number of channels, respectively. The latter further transforms the F into a categorical logit map $Z \in \mathbb{R}^{H\times W\times 2}$. The conventional detection task loss is to train each pixel with its ground-truth label using cross-entropy:

$$\mathcal{L}_{TASK} = \frac{1}{N}\sum_{h,w} CE\left(p_{I_{db}}^{(h,w)}, y_{I_{db}}^{(h,w)}\right), \tag{1}$$

here, $N = H \times W$, denotes the normalization constant. CE denotes the cross-entropy loss. $p_{I_{db}}^{(h,w)} = \sigma(Z_{h,w})$, where σ denotes the softmax function. $y_{I_{db}}^{(h,w)} \in \{0,1\}$ denotes the ground-truth label of the (h,w)-th pixel.

Revisiting low-light enhancement. Since this work handles bi-temporal remote sensing image pairs rather than continuous video frames, we focus this formulation on low-light image enhancement (LLIE) and its limitations for change detection under inconsistent illumination. LLIE approaches focus on enhancing the visibility and clarity of individual low-light images. Given a low-light image I_{low}, a learning-based deep network $\mathcal{F}$ with learnable parameters θ is trained to obtain an enhancement image I_{high}, and the formula could be expressed as:

$$I_{high} = \mathcal{F}(I_{low}, \theta) \tag{2}$$

However, conventional LLIE usually enhances each image independently and does not explicitly control whether two temporal images remain photometrically comparable after enhancement. In contrast, CPEN treats the two temporal images as a coupled input and learns a pairwise enhancement mapping for change detection:

$$\{I_{cb}^1, I_{cb}^2\} = \mathcal{F}(\{I_{db}^1, I_{db}^2\}, \theta), \tag{3}$$

where the enhanced pair is expected to improve scene visibility while reducing cross-temporal brightness discrepancy.

The goal of multi-brightness incremental change detection is to update the model using only images from the new brightness domain $\mathcal{L}_t$. To improve performance in the new bright domain while preserving effectiveness in the old bright domain, three strategies can be explored. One approach is to create a memory pool to store representative samples from previous data. However, it can be difficult to accurately assess the uniqueness of image pairs. Furthermore, privacy and security concerns may limit the storage of certain types of data, complicating the training process. The second option involves training separate parameter networks for specific brightness conditions, which can reduce the model's generalization capability and restrict the applicability of a single model. The third option aims to minimize the differences and offsets between various brightness domains while maintaining consistent photometric properties across diverse lighting conditions, making it more suitable for our specific problem. And this paper primarily focuses on this third strategy as the chosen approach.

$$\min_{\mathcal{F}(I_{db})}, \mathcal{L}_{TASK} = \frac{1}{N}\sum_{h,w} CE\left(p_{I_{db}}^{(h,w)}, y_{I_{db}}^{(h,w)}\right), \tag{4}$$

Consequently, the central component of our proposed change detection framework is designed to enhance low images captured at different times while reducing the disparity in brightness distribution between the input image pairs.

3.2. Proposed Framework

To tackle multi-brightness change detection learning, we propose the Consistent Photometric Enhancement Network (CPEN). Our proposed method begins with the low-light compensation mechanism module, which is designed to reconstruct consistent dark light $\widetilde{I_{cb}}$ for original input bi-temporal images I_{db} with fully learning low-frequency signals and suppressing high-frequency signals. This module could standardize the level of degradation, resulting in a more uniform degradation distribution that helps alleviate the complexity of subsequent tasks. Following this, we enhance the consistency degradation of

bi-temporal images by constraining the distribution and preserving the shared features in different brightness domains. This enhanced module could effectively improve bi-temporal images' quality to obtain high brightness images I_{cb}. In the end, the change detection module is used to extract features of every pixel from the enhanced images and determine where changes $p_{I_{db}}^{(h,w)}$ have occurred.

This multi-stage but end-to-end approach enhances change detection accuracy and robustness while minimizing the differences and offsets between various brightness domains and improving the bi-temporal images' quality. An overview of the complete pipeline can be found in Figure 2.

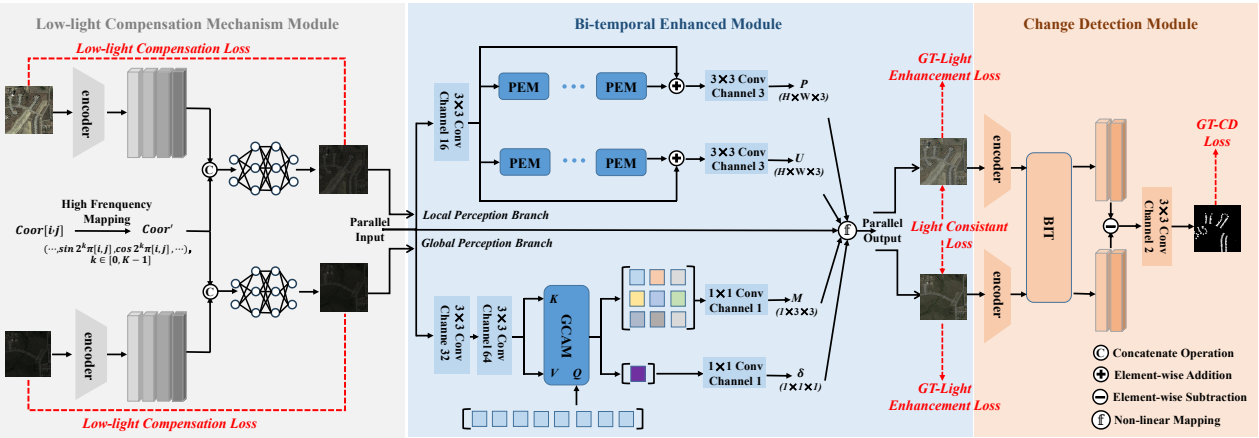


Figure 2. The pipeline of our proposed consistent photometric enhancement network (CPEN), which consists of three basic modules: low-light compensation mechanism module, bi-temporal enhanced module, and change detection module.

3.3. Low-Light Compensation Mechanism Module

In the low-light compensation mechanism module (LLCMM), two distinct branches of a Siamese compensation network independently process bi-temporal images while sharing parameters between them. The fundamental concept is to employ normalized degradation levels alongside neural representations (NR) to achieve a more uniform degradation distribution. Concretely, in NR, the image I_{db} is converted into a feature map $E_l \in \mathbb{R}^{H_l \times W_l \times C_l}$. Here, H_l and W_l denote the spatial height and width of the LLCMM feature, and C_l denotes the channel number of the encoded low-light feature. Meanwhile, the position of each pixel is documented in a coordinate set $Coor \in \mathbb{R}^{H_l \times W_l \times 2}$, where the 2 denotes the dimension of horizontal and vertical coordinates. The image I_{db} can be characterized by its features along with a set of coordinates. In the low-light compensation mechanism module illustrated in the left part of Figure 2, we combine $Coor$ and E_l, then apply a decoding function F_{MLP} to produce the output image $\widetilde{I_{cb}}$. The low-light compensation procedure of the image is formulated as:

$$\widetilde{I_{cb}}[i,j] = F_{MLP}(E_l[i,j], Coor[i,j]) \tag{5}$$

where the pixel's location is denoted by $[i,j]$ $(1 \le i \le H_l,\ 1 \le j \le W_l)$, with $\widetilde{I_{cb}}[i,j]$ representing the resulting RGB value. By forecasting the RGB values for each pixel, a low-light compensation image is reconstructed. We promote the similarity between $\widetilde{I_{cb}}[i,j]$ and I_{db} using the ℓ_1-norm. This loss does not aim to enhance illumination, but to normalize degradation patterns across bi-temporal images by aligning their low-frequency photometric components. This low-light compensation loss related to neural representations is articulated as:

$$\mathcal{L}_{lc} = ||\widetilde{I_{cb}} - I_{db}|| \tag{6}$$

Using the trained F_{MLP} function, every feature map E_l has the ability to obtain the image neural representation. This function maps coordinates to their corresponding predicted RGB values. The F_{MLP} relies on E_l to represent a range of RGB values accurately at the same coordinates *Coor*. Conversely, without *Coor*, the E_l cannot achieve normalization of degradation by adapting its fitting capability.

Neural networks typically emphasize lower-frequency information. While our decoding function can estimate RGB values, it may overlook some high-frequency details during the rendering process. For instance, neighboring pixels along an edge may exhibit significant RGB variations despite minor changes in coordinates. This implies that the F_{MLP} generates distinct outputs with similar coordinate inputs, which is a challenge. Drawing inspiration from reference [43], to better accommodate high-frequency variations, we introduce positional encoding by mapping input coordinates to a higher-dimensional space before feeding them into the F_{MLP}. This process, depicted in the left area of Figure 2, involves utilizing a high-frequency function to transform the original coordinates *Coor* from $\mathbb{R}$ into a higher space $\mathbb{R}^{2L'}$, as expressed below:

$$Hl = (\ldots, sin(2^i \pi[i,j]), cos(2^i \pi[i,j]), \ldots) \tag{7}$$

where i is set from 0 to $L' - 1$, L' represents a hyperparameter determining the dimension value; the higher-dimensional coordinates are constructed as: $coor' = Hl(coor)$. It is important to recognize that adjusting the value of L' enables us to modify the fitting capacity of our low-light compensation mechanism module. Specifically, a larger L' leads to a more accurate fit; to balance the performance and efficiency, L' is set to 4.

The physical meaning of LLCMM can be summarized as low-frequency photometric normalization. For an unchanged/shared region Ω_u, the optical imaging model gives

$$I^t(x) = R^t(x) \odot L^t(x) + n^t(x), \quad x \in \Omega_u : \Delta I(x) \approx R(x) \odot \Delta L(x) + \Delta n(x), \tag{8}$$

where R^t, L^t, and n^t denote surface reflectance, exposure, and noise, respectively. This indicates that pseudo temporal differences in unchanged regions mainly come from ΔL rather than real land-cover changes. Therefore, LLCMM aims to reduce the low-frequency luminance discrepancy:

$$\begin{aligned} \mathcal{D}_{lf}(\widetilde{I}^1_{cb}, \widetilde{I}^2_{cb}) &\leq \mathcal{D}_{lf}(I^1_{db}, I^2_{db}), \\ \mathcal{D}_{lf}(A, B) &= \|G_\sigma * Y(A) - G_\sigma * Y(B)\|_1, \end{aligned} \tag{9}$$

where $Y(\cdot)$ denotes the luminance component and G_σ denotes a low-pass filter. Thus, LLCMM provides a more comparable low-frequency photometric basis without forcing structural or semantic changes to disappear.

3.4. Bi-Temporal Enhanced Module

For the bi-temporal images, the consistency degradation of bi-temporal images $\widetilde{I_{cb}}$ is obtained with a low-light compensation mechanism module. To further enhance the light brightness of low-light images and improve the quality of bi-temporal images, a bi-temporal enhanced module (BTEM) is designed, consisting of a local perception branch and a global perception branch, which is described in the middle of Figure 2. For processing the degraded bi-temporal images $\widetilde{I_{cb}}$ for consistency, the enhanced module produces the target RGB image I_{cb} under a uniform and consistent lighting condition.

Local Perception Branch: To retain the original resolution within the local branch and safeguard information details, a transformer architecture has been devised for this specific

branch. This design enables the handling of images of arbitrary resolutions without the need for resizing.

Initially, we increase the channel dimension using a 3×3 convolution and feed them into two parallel branches arranged with the Pixel-wise Enhancement Module (PEM). To streamline the design within the local branch, we substitute self-attention with depth-wise convolution, which aids in reducing parameters and cutting down on computational costs. Illustrated in Figure 3, the PEM starts by encoding positional information through a 3×3 depth-wise convolution before refining local details with PWConv-DWConv-PWConv. Subsequently, we employ two 1×1 convolutions to enhance token representation individually. Notably, we introduce a new form of normalization called light normalization to replace the Layer Normalization in transformers [44]. As depicted in Figure 3, light normalization adjusts scale a and bias b through two trainable parameters before combining channels via a learnable matrix initialized as an identity matrix. To aid convergence, Layer Scale [45] is incorporated to scale features by a small factor k_1, k_2.

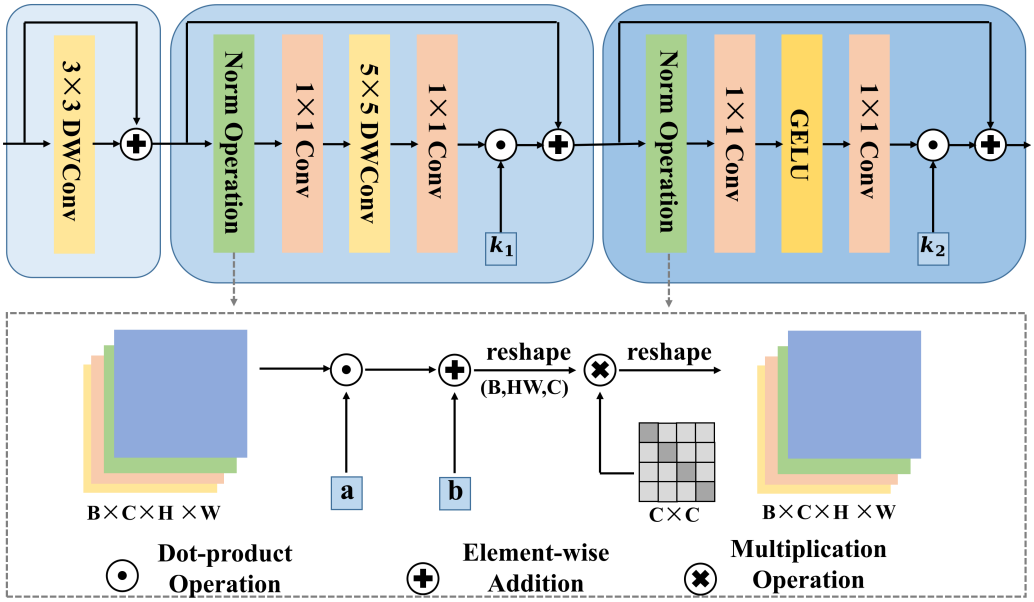


Figure 3. The details of the Pixel-wise Enhancement Module (PEM).

In the upper middle part of Figure 2, we place three PEMs one after the other and subsequently link the resulting features with the initial features via element-wise addition. The skip connections aid in preserving the original image details. Lastly, we reduce the channel dimension using 3×3 convolutions and apply the ReLU/Tanh function to produce the local components $P \in \mathbb{R}^{H \times W \times 3}$ and $U \in \mathbb{R}^{H \times W \times 3}$.

Global Perception Branch: We introduce global component queries for the decoding and prediction of $M \in \mathbb{R}^{3 \times 3}$ and $\delta \in \mathbb{R}^{1}$ in the lower middle part of Figure 2. This transformer framework facilitates capturing overarching interactions between the context and individual pixels. The approach initially involves stacking two convolutions to serve as a lightweight encoder. This encoder operates by encoding features at a heightened dimension but with reduced resolution. This strategy not only economizes computational expenses, contributing to a lightweight design, but also enhances the extraction of global-level features from the image.

Subsequently, the derived features undergo processing in the global cross-attention module (GCAM). The structure of the GCAM is depicted in Figure 4. Notably, our global component queries Q commence as zeros, eschewing the need for additional multi-head self-attention mechanisms. Q represents a learnable embedding for global components,

directing attention toward keys K and values V derived from the encoded features. The positional encoding for K and V is achieved through a depth-wise convolution, accommodating various input resolutions seamlessly. Following the feed-forward network (FFN) with two linear layers, we introduce two extra parameters with specialized initialization to yield the color matrix and gamma. This deliberate initialization ensures that the color matrix corresponds to the identity matrix M, while the initial gamma value δ is set to one, thereby fostering stable training conditions.

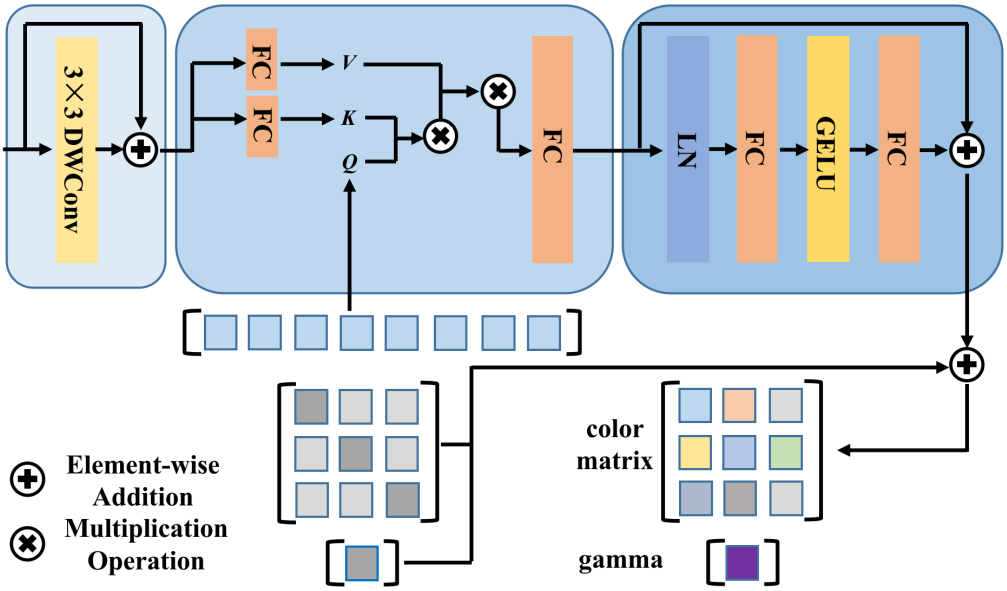


Figure 4. The details of the global cross-attention module (GCAM).

By utilizing both the local and global branches, we obtain the local components P and U, along with the global parameter identity matrix M and the gamma value δ. Subsequently, the enhanced images are derived through a non-linear mapping, which is represented as follows:

$$\mathbb{F} = (max((\widetilde{I_{cb}} \odot P + U) \otimes M), 0)^{\delta} \tag{10}$$

Following the enhanced module, we obtain I_{cb} and then promote the similarity between I_{cb} and I_{gt} using the ℓ_1-norm to construct the loss function named as GT-Light Enhancement Loss (GTLEloss). To further regulate the brightness similarity of the enhanced bi-temporal images, we also minimize the distance between I^1_{cb} and I^2_{cb} using the ℓ_1-norm, leading to the creation of a loss function called Light Consistent Loss (LCloss). Consequently, the overall loss for this part is expressed as:

$$\mathcal{L}_{lh} = ||I^1_{cb} - I^1_{gt}|| + ||I^2_{cb} - I^2_{gt}|| + ||I^1_{cb} - I^2_{cb}|| \tag{11}$$

Based on the low-frequency photometric normalization introduced by LLCMM, the Light Consistent Loss further regularizes the enhanced bi-temporal images to suppress residual illumination-induced pseudo differences. It is used as an image-level consistency regularizer rather than a constraint that forces true changed regions to become identical; true structural and semantic changes are preserved by the supervised change detection loss $\mathcal{L}_{cd}$.

3.5. Change Detection Module

In the change detection module (CDM), we adopt the CNN-Transformer-based Bi-temporal Interaction Transformer (BIT) framework [7] as the change detection backbone. The change detection model is illustrated in Figure 5. Initially, our model employs multiple

convolution blocks to extract feature maps from each input image, which are then channeled into the BIT for refining bi-temporal features. Subsequently, these refined feature maps are directed to a prediction module for generating predictions at the pixel level. A pivotal aspect of our approach is that the BIT excels at understanding and correlating global contextual information regarding high-level semantic concepts, subsequently enhancing the original bi-temporal features.

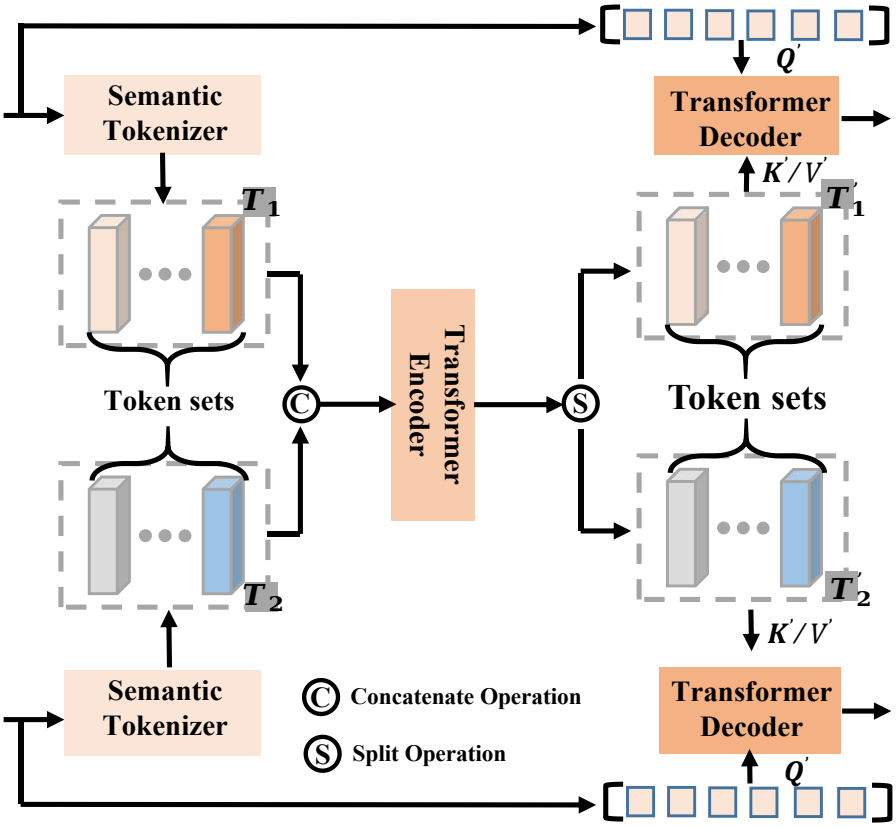


Figure 5. The details of the change detection module with bi-temporal interaction transformer (BIT).

The BIT mainly comprises three primary components: (1) A Siamese semantic tokenizer that clusters pixels into concepts to create a concise set of semantic tokens for each temporal input. (2) A transformer encoder that captures the contextual relationships among semantic concepts in a token-based space-time representation. (3) A Siamese transformer decoder that maps the associated semantic tokens back to pixel-space, refining the feature map for each temporal input in the process.

Thus, the detailed procedure is described in Figure 5. Firstly, extract the high-level feature maps X^1 and X^2 of the enhanced bi-temporal images I^1_{cb} and I^2_{cb} with a CNN backbone. Then, compute the token set for each temporal features T_1 and T_2 and concatenate them into $T = \text{Concat}(T_1, T_2)$, where Concat denotes the concatenation operation. Formally,

$$T^i = (A^i_t)^T X^i = (\sigma(\phi(X^i; W)))^T X^i \tag{12}$$

where, $\phi(.)$ denotes the point-wise convolution with a learnable kernel $W \in \mathbb{R}^{C \times L}$, C is the channel of bi-temporal images, and L is the vocabulary set size. $\sigma(.)$ is the softmax function to normalize each semantic group to obtain the attention maps $A^i_t \in \mathbb{R}^{H \times W \times L}$. Following this, use the transformer encoder to generate context-rich tokens. In the transformer encoder, the fundamental concept revolves around multi-head self-attention (MSA), where multiple independent attention heads operate simultaneously. The resulting outputs are combined and subsequently transformed to yield the ultimate values. One key benefit of

MSA lies in its capacity to collectively focus on data across various representation subspaces and positions. To elaborate formally,

$$MSA(T^{(l-1)}) = \text{Concat}(head_1, \ldots, head_h)W^o \tag{13}$$

where

$$head_j = Att(T^{(l-1)}W_j^q, T^{(l-1)}W_j^k, T^{(l-1)}W_j^v), \tag{14}$$

$$Att(Q', K', V') = \sigma(Q'K'^T/\sqrt{d})V'. \tag{15}$$

The matrices $W_j^q, W_j^k, W_j^v \in \mathbb{R}^{C\times d}$ and $W^o \in \mathbb{R}^{h\times d\times C}$ are linear projection matrices, h is the head number, and d is the channel dimension of the triple. Then use the multilayer perceptron to get the final transformer encoder representation using the GELU activation [46]: $MLP(T^{(i-1)}) = GELU(T^{(l-1)}W_1)W_2$, where $W_1 \in \mathbb{R}^{C\times 2C}$ and $W_2 \in \mathbb{R}^{2C\times C}$.

Until now, we have obtained two sets of context-rich tokens T_1' and T_2' and then used a transformer decoder to refine the original features. This transformer decoder consists of multihead cross-attention (MA) and MLP blocks. Formally, for each layer, MA is defined as follows,

$$MA(X^{i,(l-1)}, T_i') = \text{Concat}(head_1, \ldots, head_h)W^o \tag{16}$$

where $head_j = Att(X^{i,(l-1)}W_j^q, T_i'W_j^k, T_i'W_j^v)$. After extracting high-level semantic features with a CNN backbone and a BIT module, we obtain two upsampled feature maps X^{1*} and X^{2*}, and then a fully connected network is employed for the last change discrimination. The predicted change probability map is given by,

$$p_{I_{db}}^{(h,w)} = \sigma(g(|X^{1*} - X^{2*}|)) \tag{17}$$

input the element-wise absolute of the subtraction of the two feature maps into the last change classifier, and σ denotes a softmax function on a per-pixel basis along the channel dimension of the classifier's output.

During the training procedure, we aim to minimize the cross-entropy loss [42] to refine the network parameters; the loss function is formally expressed as:

$$\mathcal{L}_{cd} = 1/(H \times W)\sum^{h,w} l(p^{(h,w)}, y^{(h,w)}) \tag{18}$$

where $l(p^{(h,w)}, y^{(h,w)}) = -log(p^{(h,w)}, y^{(h,w)})$ is the cross-entropy loss, and $y^{(h,w)}$ is the ground-truth label.

3.6. Optimization Strategy

To sum up, the parameters of CPEN can be optimized through the overall loss function below:

$$\mathcal{L}_{all} = \lambda_1 * \mathcal{L}_{lc} + \lambda_2 * \mathcal{L}_{lh} + \lambda_3 * \mathcal{L}_{cd} \tag{19}$$

where λ_1, λ_2 and λ_3 are the hyperparameters for the Consistent Photometric Enhancement Network by balancing the contribution of different modules during the training step.

The values of λ_1, λ_2, and λ_3 are determined according to the functional dependency and convergence order of the three objectives, rather than by an arbitrary fixed weighting. Specifically, the compensation objective $\mathcal{L}_{lc}$ should dominate the early optimization because LLCMM first reduces the cross-temporal photometric discrepancy and provides a stable low-light basis for the following enhancement stage. After the compensation branch becomes stable, the enhancement objective $\mathcal{L}_{lh}$ is assigned the largest weight so that BTEM can learn brightness restoration and bi-temporal photometric consistency on normalized

inputs. In the final stage, the detection objective $\mathcal{L}_{cd}$ is emphasized because the enhanced images are sufficiently reliable for discriminative change representation learning. Therefore, the adopted weights follow a progressive schedule from photometric normalization, to enhancement consistency, and finally to change discrimination.

At different training stages, the overall network emphasizes different sub-modules, as summarized in Algorithm 1. This stage-wise optimization strategy is designed to stabilize the joint learning of three tightly coupled objectives with different convergence characteristics. In the first stage ($epoch \leq N^1_{epoch}$, where $N^1_{epoch} = 1000$), the Low-light Compensation Mechanism Module is assigned the highest priority, since it establishes a more consistent photometric basis for the bi-temporal image pair and facilitates the subsequent optimization of the Bi-temporal Enhancement Module and the Change Detection Module. Accordingly, we set $\lambda_1 = 0.6$, $\lambda_2 = 0.3$, and $\lambda_3 = 0.1$. In the second stage ($epoch \leq N^2_{epoch}$, where $N^2_{epoch} = 10{,}000$), the training focus shifts to the Bi-temporal Enhancement Module, because once the compensation branch becomes sufficiently stable, the network can more reliably learn brightness restoration and shared structural representations. To balance the contributions of the three sub-modules at this stage, we set $\lambda_1 = 0.1$, $\lambda_2 = 0.7$, and $\lambda_3 = 0.2$. In the final stage ($epoch > N^2_{epoch}$), the emphasis is further transferred to the Change Detection Module, as the enhanced inputs become more suitable for discriminative semantic learning. Therefore, the loss weights are adjusted to $\lambda_1 = 0.1$, $\lambda_2 = 0.2$, and $\lambda_3 = 0.7$. We emphasize that this schedule is a practical strategy for stabilizing optimization rather than a theoretically unique solution. To further examine its robustness, we added additional analyses in the following experiments by varying both the stage transition thresholds and the loss weights. The results show that the proposed framework remains relatively stable within a reasonable range of settings.

Algorithm 1 Training process in the t_{th} step in different stages.

Input:
$(I^1_{db}, I^2_{db}, I^1_{gt}, I^2_{gt}, y^{(h,w)})$: New bi-temporal images and corresponding labels for the current t-step
Output:
$p^{(h,w)}$ The predicted change probability map.

```
Initialize Low-light Compensation Mechanism Module, Bi-temporal Enhanced Module and Change Detection Module.
for epoch do
  if epoch ≤ N^1_epoch then
    In the first training stage, set λ1 = 0.6, λ2 = 0.3, λ3 = 0.1, and focus the network on optimizing the Low-light Compensation Mechanism Module.
  else if epoch ≤ N^2_epoch then
    In the second training stage, set λ1 = 0.1, λ2 = 0.7, λ3 = 0.2, and focus the network on optimizing Bi-temporal Enhanced Module.
  else
    In the third training stage, set λ1 = 0.1, λ2 = 0.2, λ3 = 0.7, and focus the network on optimizing Change Detection Module.
  end if
  Update the weights of the focused module mainly.
end for
```

4. Results

4.1. Experimental Setup

The experiments are organized with two different dataset roles. NPULL-CD is used as the real-world validation dataset because it contains remote sensing image pairs acquired under daytime and nightfall conditions. The gamma-based versions of LEVIR-CD, PRCV-

CD, and SYSU-CD are used as controlled stress-test datasets, where the brightness degradation level can be adjusted while the original geometry and annotations remain unchanged. Therefore, the simulated datasets are not treated as a replacement for real low-light acquisition, but as a way to test method robustness under known illumination perturbations.

4.1.1. Simulated Datasets

The simulated datasets are designed for controlled robustness analysis rather than for representing all real-world low-light imaging effects. Given the lack of a pre-existing multi-brightness change detection dataset, we compiled three datasets tailored for multi-brightness change detection learning, drawing from the following three widely used experimental datasets. The examples are shown in Figure 6, and the details are provided below. To simulate varying illumination conditions, image brightness is adjusted using gamma correction, which applies a nonlinear power-law transformation to pixel intensities. The gamma correction process can be formulated as

$$I_{\text{out}} = I_{\text{in}}^{1/\gamma}, \tag{20}$$

where I_{in} and I_{out} denote the normalized input and output pixel intensities, respectively, and γ is the gamma parameter controlling the degree of brightness adjustment. In this study, gamma values in the range of 0.2–2 are used to intentionally darken the images, thereby generating low-light scenarios with different illumination levels.

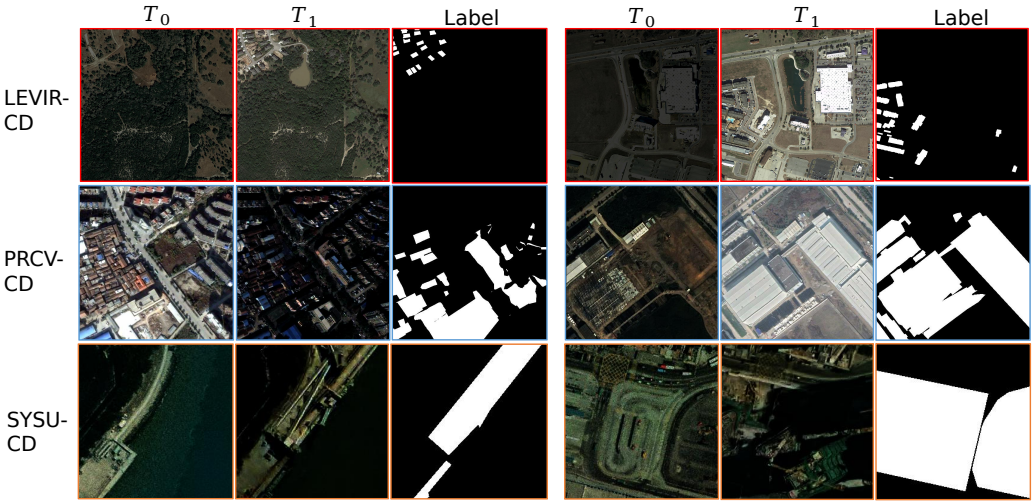


Figure 6. Examples of simulated change detection datasets. The simulated data are generated from existing remote sensing change detection datasets by applying a global gamma transformation to construct controllable low-light stress-test conditions.

The adoption of gamma correction is motivated by its controllability: it allows us to isolate the effect of radiometric degradation and construct illumination stress tests with different severity levels. Nevertheless, the global gamma transformation cannot fully reproduce real low-light remote sensing acquisition, where illumination changes may be spatially nonuniform and coupled with shadows, exposure differences, sensor noise, and imaging-pipeline distortions. To quantify this discrepancy, we further analyze sampled subsets from the three simulated datasets and the real-world NPULL-CD dataset. The statistics are summarized in Table 1.

Table 1. Illumination-discrepancy statistics computed from sampled subsets of the simulated and real-world datasets. Mean luminance and dark-pixel ratio are reported for the two temporal images A/B. Luminance values are normalized to $[0, 1]$, and dark pixels are defined as pixels with luminance $Y < 0.2$.

Dataset	Type	Mean Y (A/B)	Dark Pixels (A/B, %)	Pair Gap	JS Div.	Local Std.
LEVIR-CD	Simulated	0.276/0.055	57.94/93.15	0.246	0.266	0.081
PRCV-CD	Simulated	0.234/0.109	68.35/84.00	0.171	0.242	0.078
SYSU-CD	Simulated	0.128/0.120	87.95/84.95	0.055	0.185	0.038
NPULL-CD	Real-world	0.599/0.314	0.32/24.34	0.285	0.422	0.107

Note: Pair gap denotes the average absolute difference between the image-level mean luminance of paired A/B images. JS div. denotes the average Jensen–Shannon divergence between paired luminance histograms. Local std. denotes the average standard deviation of 16×16 block-wise mean luminance and reflects spatial illumination nonuniformity.

As shown in Table 1, the gamma-simulated datasets provide controllable underexposed samples, as reflected by their high dark-pixel ratios. However, the real-world NPULL-CD subset exhibits a larger paired histogram discrepancy (JS divergence of 0.422) and stronger spatial illumination nonuniformity (local std. of 0.107). These results indicate that gamma simulation is useful for controlled robustness evaluation, but it remains a simplified approximation of real illumination degradation. Therefore, in this work, the simulated datasets are used as stress-test data, while NPULL-CD is used to evaluate the proposed method under real illumination inconsistency.

LEVIR-CD: The LEVIR-CD dataset [23] is a publicly available extensive building change detection dataset. It comprises 637 pairs of high-resolution (0.5 m) remote sensing images sized at 1024×1024. We adhere to its default dataset division (training/validation/test) and introduce random brightness adjustments to the images. Due to GPU memory constraints, we segment the images into non-overlapping small patches of size 256×256.

PRCV-CD: The PRCV dataset, derived from multi-temporal remote sensing images of major first-tier cities and the SECOND [47] dataset, comprises 10,000 pairs of 512×512 images, and we further divide these images into non-overlapping small patches of size 256×256 in this experimental part. Predominantly covering cities like Beijing, Shanghai, Guangzhou, and Hangzhou, the dataset focuses on changes in ground objects, particularly buildings.

SYSU-CD: The SYSU-CD dataset [48] consists of 20,000 pairs of aerial images with a spatial resolution of 0.5 m, collected over Hong Kong between 2007 and 2014. The dataset covers a wide range of land-cover changes, including urban construction, suburban expansion, and large-scale infrastructure development. Owing to the complex urban morphology and rapid development of the region, SYSU-CD exhibits high diversity and structural complexity, making it a challenging benchmark for change detection tasks. All image pairs are cropped into patches of size 256×256 pixels for model training and evaluation.

4.1.2. Realistic Dataset

To evaluate the proposed method under practical illumination inconsistency, we use NPULL-CD as the real-world validation dataset. Additionally, we constructed a real-world change detection dataset collected over the Chang'an Campus of the Northwestern Polytechnical University, called the Northwestern Polytechnical University Low-Light Change Dataset (NPULL-CD). The sequence data were acquired at two different time points: daytime on 31 March 2019, and at nightfall on 14 April 2026, covering an area of approximately 1.7 km $\times$ 1.3 km. The dataset generation process is illustrated in Figure 7. First, data from each time period were processed to generate corresponding maps. These

maps were then divided into tiles, followed by feature-based alignment to ensure spatial consistency between the two temporal datasets. Finally, we obtained a set of spatially aligned paired tiles representing different times. In total, 299 pairs of 256 × 256 image tiles were generated, which were split into training and testing sets with a ratio of 8:2.

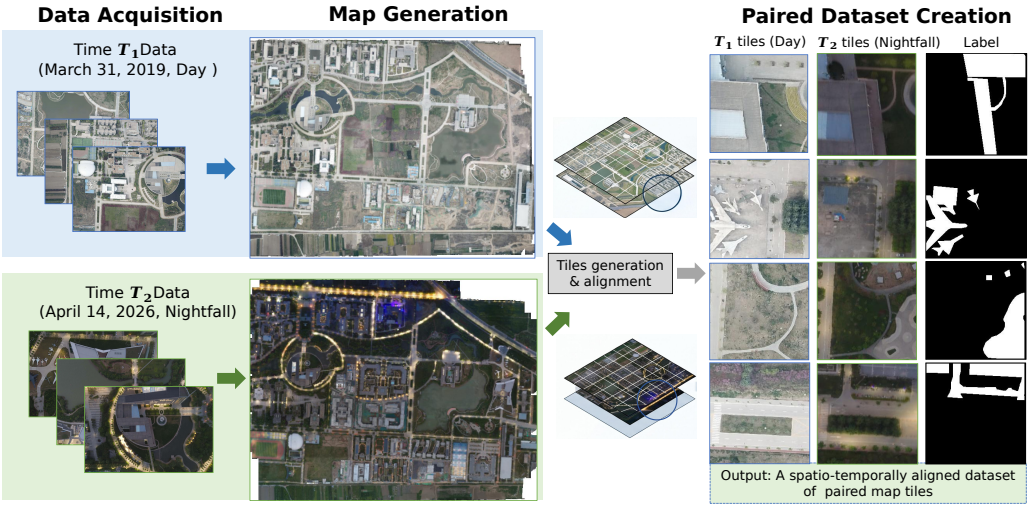


Figure 7. Examples of realistic change detection datasets. The dataset is constructed through a pipeline including data acquisition, map generation, tile generation, and alignment.

The NPULL-CD tiles are organized as zoom-level-19 Web Mercator tiles. The estimated original ground sampling distance is approximately 0.247 m/pixel, and each 256 × 256 original tile covers about 63.35 m on the ground. During co-registration, image pairs are first matched by file-name keys. The nightfall image is then aligned to the daytime reference image using a gradient-based translation search: Gaussian-smoothed grayscale images are converted to Sobel gradient-magnitude maps, a B-to-A integer translation is searched within ±96 pixels, and the best shift is selected by normalized gradient correlation with an overlap constraint. The aligned image is warped using a 2-D translation affine transform, the common valid region is cropped, and the result is resized to 256 × 256 using area interpolation. After this common-region crop and resizing, the average sampling interval is about 0.212 m/pixel horizontally and 0.218 m/pixel vertically, which should be understood as the post-resampling pixel interval rather than improved native image resolution.

Because independent manually surveyed ground control points are not available, we report automatic proxy checks for the co-registration accuracy. Based on stable SIFT matches after applying the estimated translation, 72 check points from 45 usable image pairs give a mean residual error of 1.90 pixels (0.47 m), an RMSE of 2.16 pixels (0.53 m), and a maximum error of 3.97 pixels (0.98 m). A more conservative local NCC grid check on the final aligned tiles uses 604 check points from 105 usable image pairs and gives a mean residual error of 2.12 pixels (0.46 m), an RMSE of 2.61 pixels (0.56 m), and a maximum error of 5.66 pixels (1.22 m). These values indicate that small residual misregistration may still remain near object boundaries and can introduce false detections, but the image pairs are sufficiently aligned for evaluating low-light change detection.

4.1.3. Evaluation Metrics

Change detection performance is commonly assessed using four key metrics: precision, recall, F1 score, and intersection over union (IoU). Precision gauges the model's false positive rate, while recall assesses the false negative rate. The F1 score combines both precision and recall, with a higher F1 score indicating superior model performance. IoU quantifies the overlap between the change category on the detection map and the corre-

sponding ground truth, highlighting their agreement level. Typically, an IoU exceeding 0.5 signifies a favorable outcome. The definitions of these metrics are as follows:

$$\text{precision} = \frac{\text{TP}}{\text{TP} + \text{FP}}, \tag{21}$$

$$\text{recall} = \frac{\text{TP}}{\text{TP} + \text{FN}}, \tag{22}$$

$$\text{F1} = \frac{2\text{precision} \cdot \text{recall}}{\text{precision} + \text{recall}}, \tag{23}$$

$$\text{IoU} = \frac{|\text{Predicted-binary-masks} \cap \text{Ground-truth}|}{|\text{Predicted-binary-masks} \cup \text{Ground-truth}|}, \tag{24}$$

where TP, FP, TN, and FN refer to true positives, false positives, true negatives, and false negatives, respectively, and $|\cdot|$ denotes the number of pixels in the corresponding region.

4.1.4. Implementation Details

We implemented our models using the PyTorch 2.7.1 deep learning framework and trained them on two RTX 4090 GPUs, each equipped with 24 GB of memory. During training, standard data augmentation strategies—including flipping, rescaling, cropping, and Gaussian blurring—were applied to the input image patches. Model optimization was performed using the Adam optimizer with a learning rate of 0.001, a momentum of 0.95, and a weight decay of 0.005. To ensure efficient computation, the batch size was adapted to the spatial resolution of the datasets. For datasets with an input size of 256×256 pixels, we set the batch size to 16 to achieve optimal performance.

4.2. Comparison Experiments

To validate the effectiveness of CPEN in remote sensing change detection, we compare it with several state-of-the-art methods, including the CNN-based models DMINet [25] and SEIFNet [26], the Transformer-based model ASCNet [27], and the Mamba-based model CDMamba [9]. For a fairer comparison, we further introduce enhancement-assisted baselines by applying representative low-light enhancement methods before the competing detectors. Specifically, four "enhancement + detection" variants are considered: RetinexFormer [14] + DMINet, RetinexFormer [14] + SEIFNet, CIDNet [49] + ASCNet, and CIDNet [49] + CDMamba. In these baselines, the bi-temporal images are first enhanced and then fed into the corresponding change detection model. This allows us to evaluate CPEN against both conventional change detection methods and stronger low-light-aware comparison pipelines.

Quantitative Analysis. To assess the robustness and generalization ability of the proposed method, we conduct extensive experiments on four benchmark datasets covering heterogeneous remote sensing conditions, including different temporal intervals, spatial resolutions, and geographic environments. The main quantitative comparisons in this subsection are summarized in Table 2, which includes both direct change detection baselines and enhancement-assisted baselines. As shown in the table, the proposed CPEN consistently achieves superior performance in terms of both F1 and IoU compared with the competing methods. The advantage is also clear on the real-world NPULL-CD dataset, where the two temporal images are affected by daytime–nightfall illumination differences, local shadows, low-light noise, and small residual registration errors. CPEN obtains 86.26% F1 and 79.02% IoU on this dataset. Among the direct change detection baselines, the best F1 is 64.21% from ASCNet and the best IoU is 56.92% from CDMamba. Among the enhancement-assisted baselines, the best F1 is 70.12% from CIDNet + ASCNet, while the best IoU is 61.08% from CIDNet+CDMamba. Therefore, CPEN improves the best enhancement-assisted results by 16.14% in F1 and 17.94% in IoU. This comparison

indicates that simply enhancing the two images separately is still insufficient for NPULL-CD, because independent enhancement may improve visibility but can also keep or even enlarge the photometric gap between daytime and nightfall images. By learning pairwise compensation before enhancement, CPEN makes unchanged regions more comparable and preserves useful change cues under real illumination inconsistency. For analyzing the NPULL-CD, the geometric alignment of the real image pairs should also be considered. As described in the dataset subsection, the nightfall image is aligned to the daytime reference through a gradient-based translation search, affine warping, common-region cropping, and resizing. Although this process reduces the dominant displacement between the two acquisitions, small residual misalignments may remain, especially near building edges, thin structures, and shadowed regions. These residual shifts can generate paired false positive and false negative pixels along object boundaries and thus lower the reported F1 and IoU. Therefore, the NPULL-CD results should be viewed as performance under practical illumination inconsistency with moderate residual geometric errors, rather than under an ideally co-registered laboratory setting. For enhancement-assisted baselines, although applying standalone low-light enhancement before change detection can improve image visibility to some extent, these cascaded "enhancement + detection" schemes still fail to match the performance of CPEN. This result indicates that the performance gain of CPEN does not simply come from brightness enhancement but also from the joint optimization of low-light compensation, bi-temporal enhancement, and change detection within a unified framework. In particular, CPEN explicitly promotes photometric consistency between the two temporal inputs while preserving discriminative change cues, whereas independent enhancement may introduce additional temporal discrepancies and limit the reliability of downstream feature extraction. This advantage becomes more evident under severe illumination discrepancies, where conventional methods often struggle to extract stable and informative features for accurate change detection.

Table 2. Quantitative comparisons. The compared methods are recent state-of-the-art methods and enhancement-assisted baselines, including DMINet, SEIFNet, ASCNet, CDMamba, RetinexFormer + DMINet, RetinexFormer + SEIFNet, CIDNet + ASCNet, and CIDNet + CDMamba. The parameters (M), FLOPs (G), and FPS of computational analysis are listed to show the computation efficiency.

Methods	Computational Analysis			LEVIR-CD		PRCV-CD		SYSU-CD		NPULL-CD	
	Params (M)	FLOPs (G)	FPS	F1 ↑	IoU ↑	F1 ↑	IoU ↑	F1 ↑	IoU ↑	F1 ↑	IoU ↑
$DMINet_{23}$ [25]	6.24	14.42	121	63.14	52.87	59.84	50.62	53.43	40.94	61.97	54.38
$SEIFNet_{24}$ [26]	27.91	**8.37**	**150**	61.83	51.66	57.29	45.15	52.74	39.43	60.19	54.05
$ASCNet_{25}$ [27]	5.98	8.49	139	64.72	52.48	60.51	48.22	54.67	41.39	64.21	55.83
$CDMamba_{25}$ [9]	11.90	49.26	37	61.33	49.59	60.48	49.87	52.04	43.06	63.06	56.92
$RetinexFormer_{23}$ [14] + $DMINet_{23}$ [25]	7.85	29.99	49	85.26	72.81	75.39	62.68	68.72	61.05	68.47	59.26
$RetinexFormer_{23}$ [14] + $SEIFNet_{24}$ [26]	29.52	23.94	92	84.39	69.48	74.57	62.78	68.44	59.47	67.41	59.23
$CIDNet_{25}$ [49] + $ASCNet_{25}$ [27]	7.86	16.06	87	85.91	73.16	76.11	64.17	71.16	60.83	70.12	60.22
$CIDNet_{25}$ [49] + $CDMamba_{25}$ [9]	13.78	56.83	32	86.74	76.02	77.88	65.83	70.89	61.96	68.95	61.08
CPEN (Ours)	**4.84**	12.07	125	**90.39**	**82.67**	**80.71**	**70.58**	**78.34**	**64.26**	**86.26**	**79.02**

Note: Best results are highlighted in **bold**, the subscript following each method denotes the publication year of the corresponding article. ↑ indicates that higher values are better.

Qualitative Analysis. The qualitative comparisons are shown in Figures 8 and 9. Overall, CPEN produces more accurate change boundaries and more reliable change patterns than the competing methods under complex illumination conditions, which is consistent with the quantitative results. As shown in Figure 8, conventional change detection methods may perform well under normal illumination, but their robustness degrades significantly when facing severe low-light degradation or large bi-temporal brightness discrepancies. This is mainly because low illumination weakens discriminative texture representation, while cross-temporal brightness inconsistency causes evident feature distribution shifts.

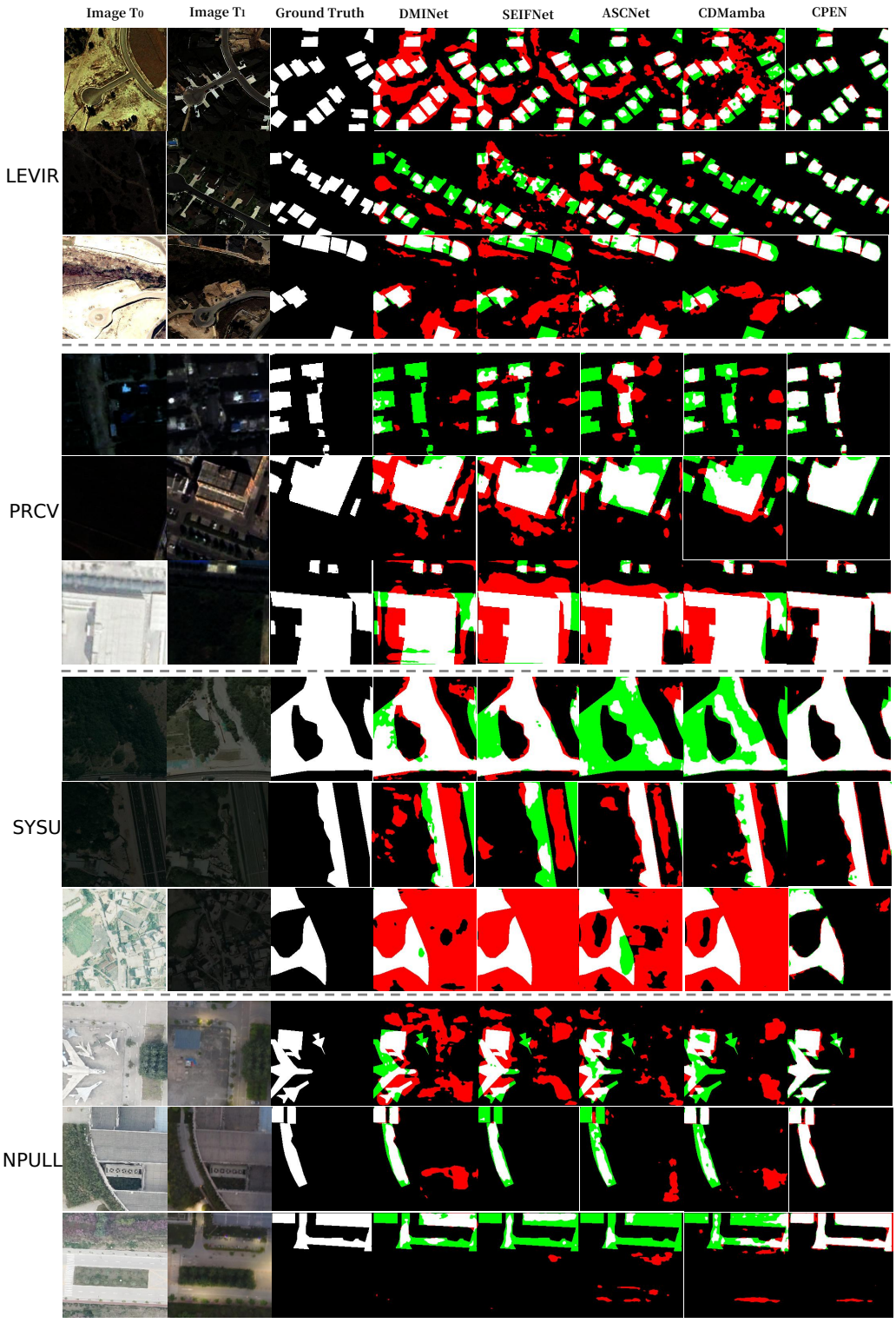


Figure 8. Qualitative comparison of change detection results on the LEVIR, PRCV, SYSU, and NPULL datasets. Results produced by the proposed method and existing state-of-the-art approaches, including DMINet, SEIFNet, ASCNet, and CDMamba, are shown on representative samples from each dataset. In the results, white denotes true positives, black denotes true negatives, red denotes false positives, and green denotes false negatives.

Figure 9 further presents the enhancement-assisted baselines and their intermediate enhancement results. When both input images are consistently dark, these methods can improve visibility and still achieve acceptable detection results, as illustrated by the examples from the LEVIR and PRCV datasets. However, when one image already has relatively high

brightness, standalone enhancement may lead to overexposure and enlarge the appearance discrepancy of unchanged regions, resulting in inferior detection performance, as observed in the SYSU and NPULL examples. In contrast, CPEN not only enhances dark regions, but also explicitly preserves photometric and feature consistency between bi-temporal images. Therefore, it avoids the inconsistency introduced by independent enhancement and achieves more stable and accurate change detection under diverse low-light scenarios.

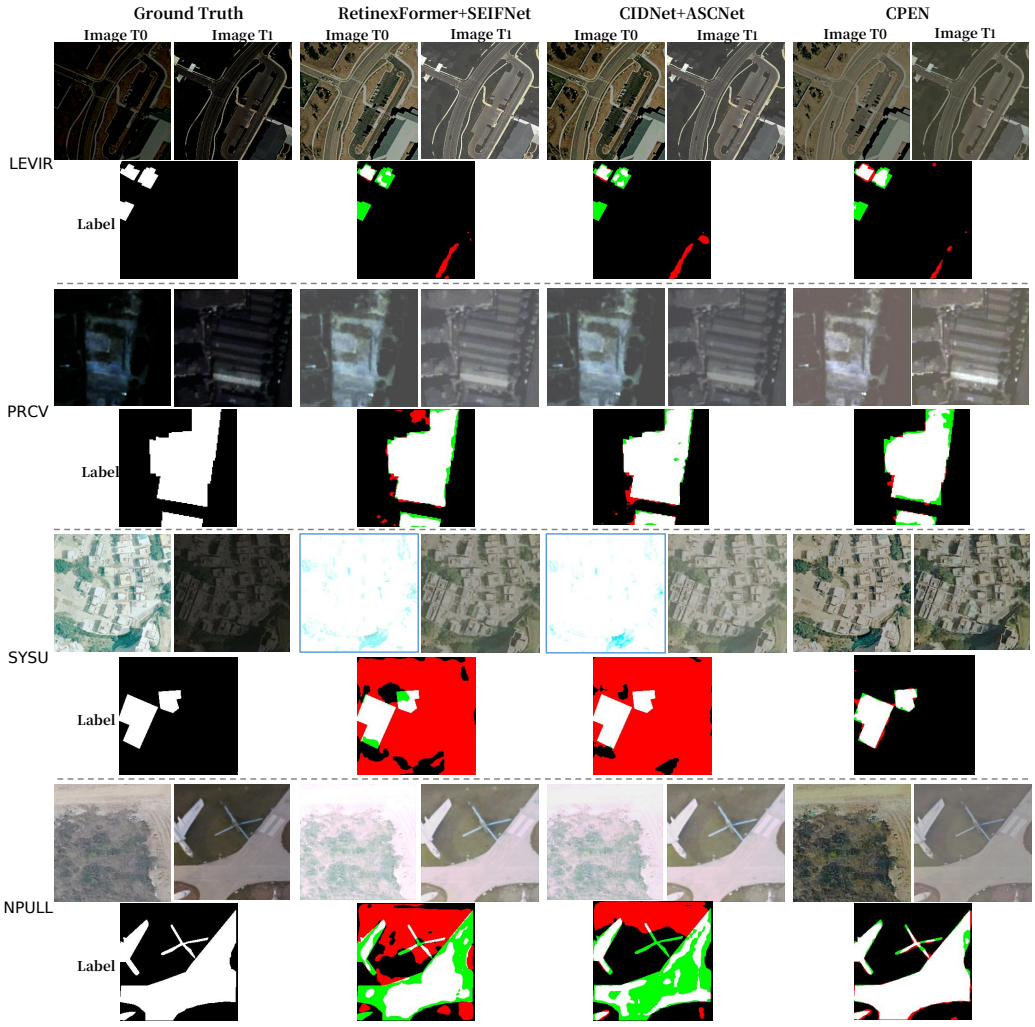


Figure 9. Qualitative comparison of change detection results on the LEVIR, PRCV, SYSU, and NPULL datasets. Results produced by the proposed method and enhancement-assisted baselines, including RetinexFormer + SEIFNet and CIDNet + ASCNet, are shown on representative samples from each dataset. In the results, white denotes true positives, black denotes true negatives, red denotes false positives, and green denotes false negatives.

4.3. Ablation Study

To validate the effectiveness of the proposed framework under complex illumination conditions, ablation studies are performed on multi-brightness change detection data. As reported in Table 3 and Figure 10, comprehensive quantitative and qualitative evaluations are carried out to analyze the individual contributions of the low-light compensation mechanism module and the bi-temporal enhanced module in the end-to-end change detection task. The results demonstrate the critical role of each component in improving robustness against illumination variations. Detailed analyses are presented in the following subsections.

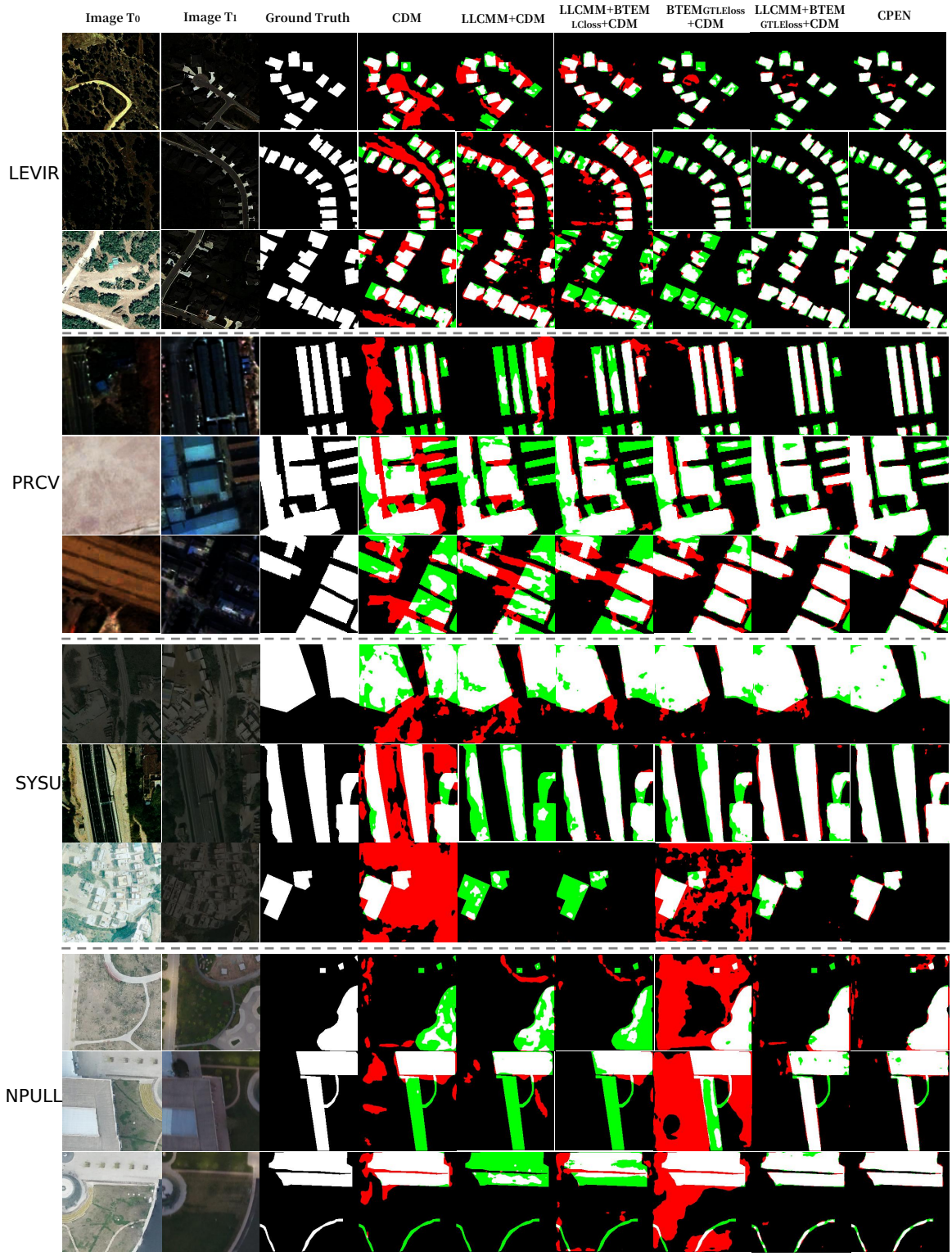


Figure 10. Qualitative visualization illustrating the effectiveness of individual modules and the role of their associated loss functions on representative samples from the LEVIR, PRCV, SYSU, and NPULL datasets. In the results, white denotes true positives, black denotes true negatives, red denotes false positives, and green denotes false negatives. Note: LLCMM = Low-light Compensation Mechanism Module; $BTEM_{LCloss}$ = Bi-temporal Enhanced Module with Light Consistent Loss; $BTEM_{GTLEloss}$ = Bi-temporal Enhanced Module with GT-Light enhancement loss; CDM = Change Detection Module.

Table 3. Quantitative ablation study results for different sub-modules of CPEN. The parameters (M), FLOPs (G), and FPS of computational analysis for different modules in CPEN are listed to show the computation efficiency.

Modules (✓ = Enabled)				Computational Analysis			LEVIR-CD		PRCV-CD		SYSU-CD		NPULL-CD	
LLCMM	$BTEM_{LCloss}$	$BTEM_{GTLEloss}$	CDM	Params (M)	FLOPs (G)	FPS	F1 ↑	IoU ↑	F1 ↑	IoU ↑	F1 ↑	IoU ↑	F1 ↑	IoU ↑
			✓	3.55	10.60	141	60.75	51.04	57.11	49.36	52.57	44.73	61.82	55.64
✓			✓	4.75	10.63	136	65.09	52.82	61.34	51.41	56.45	48.76	66.08	59.23
✓	✓		✓	4.84	12.07	125	64.58	51.37	62.51	52.69	57.50	48.32	67.12	60.95
		✓	✓	3.64	12.04	129	85.18	71.93	76.75	65.62	70.44	60.18	69.46	60.77
✓		✓	✓	4.84	12.07	125	89.79	80.24	79.43	68.55	**78.95**	63.83	85.52	77.44
✓	✓	✓	✓	4.84	12.07	125	**90.39**	**82.67**	**80.71**	**70.58**	78.34	**64.26**	**86.26**	**79.02**

Note: LLCMM = Low-light Compensation Mechanism Module; $BTEM_{LCloss}$ = Bi-temporal Enhanced Module with Light Consistent Loss; $BTEM_{GTLEloss}$ = Bi-temporal Enhanced Module with GT-Light enhancement loss; CDM = Change Detection Module. ✓ indicates that the corresponding module is enabled and the best results are highlighted in **bold** . ↑ indicates that higher values are better.

4.3.1. Study on Low-Light Compensation Mechanism Module

This section aims to systematically analyze the functionality and importance of the proposed low-light compensation module. To this end, three representative data scenarios exhibiting different levels of photometric discrepancy are selected for comparative analysis and visualization. As shown in Figure 11, scenarios with severe illumination differences present separations in their photometric histogram distributions, indicating large discrepancies in intensity distributions between the bi-temporal image pairs. Correspondingly, these photometric inconsistencies lead to substantial divergence in the feature space, which adversely affects feature alignment and downstream change detection performance.

After introducing the low-light compensation module, the photometric distributions of the compared image pairs become significantly closer, as evidenced by the reduced distance between their histogram distributions. More importantly, the feature space representations extracted from the compensated images exhibit improved alignment, demonstrating that the proposed module effectively normalizes illumination-induced degradation while preserving structural information. By narrowing both the photometric and feature space gaps between bi-temporal images, the low-light compensation module establishes a more consistent and reliable input for subsequent photometric enhancement. This consistency is critical for stable feature extraction and robust change detection under complex and varying illumination conditions.

Different from Figure 11, which analyzes histogram and feature-space alignment after low-light compensation, Figure 12 focuses on the side effect of applying independent enhancement to image pairs with mixed brightness levels. One phenomenon worth noting, as illustrated in Figure 12, is that although many existing methods are capable of enhancing low-light images individually, their application to images with normal illumination may lead to unintended degradation. Specifically, photometric enhancement models tend to excessively suppress high-intensity information in well-lit images, resulting in the loss of important visual details. In contrast, the proposed compensation-based strategy exhibits better generalization across different illumination conditions, as it does not rely on a single enhancement scheme and therefore avoids the negative effects that may arise when applying uniform enhancement to images with varying brightness levels.

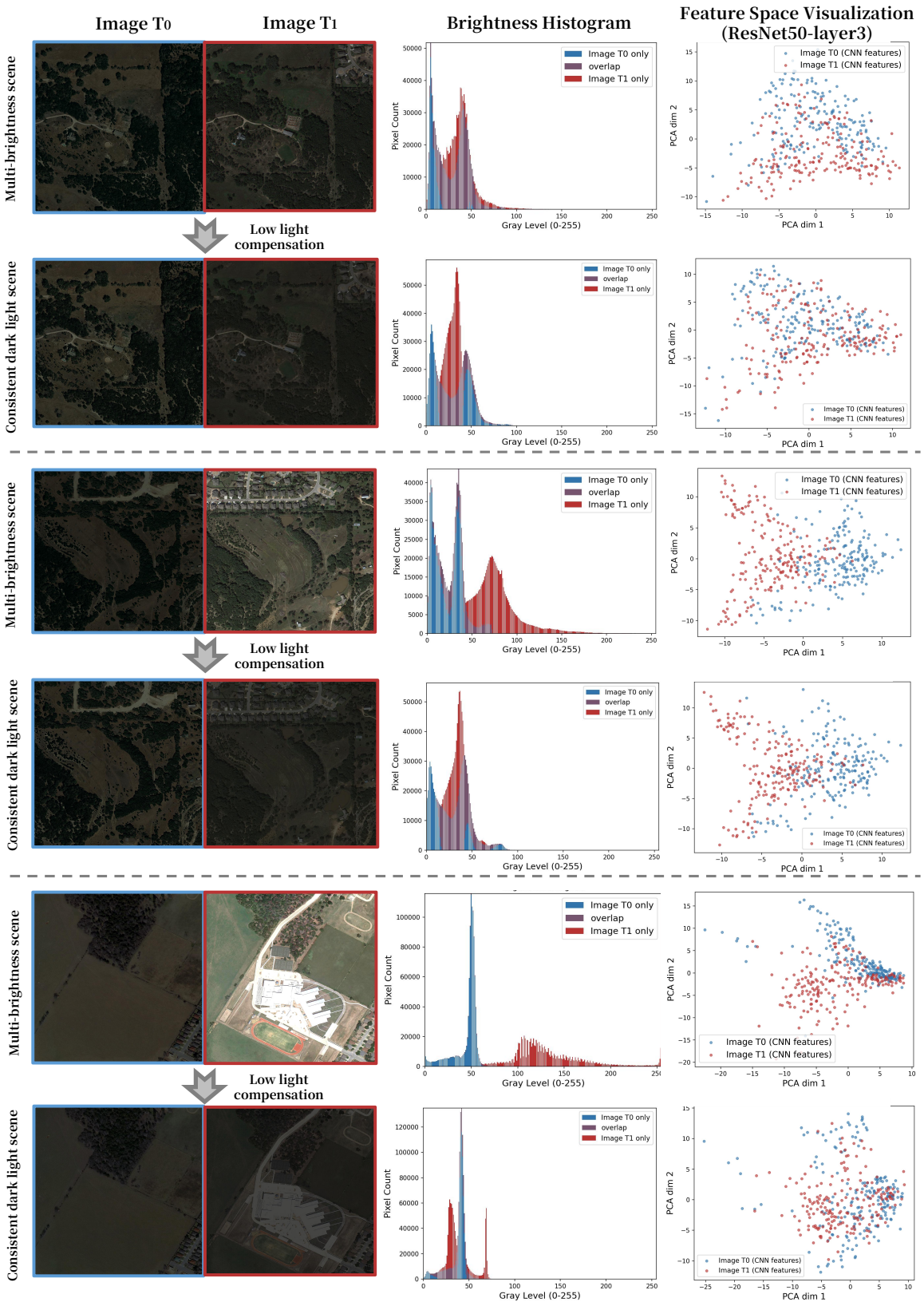


Figure 11. Visual comparisons between multi-brightness scenes and consistent dark-light scenes generated by the proposed module on the LEVIR-CD dataset (**left**). Brightness histograms and feature space visualizations (**right**) demonstrate that the proposed method effectively reduces photometric distribution discrepancies across bi-temporal images.

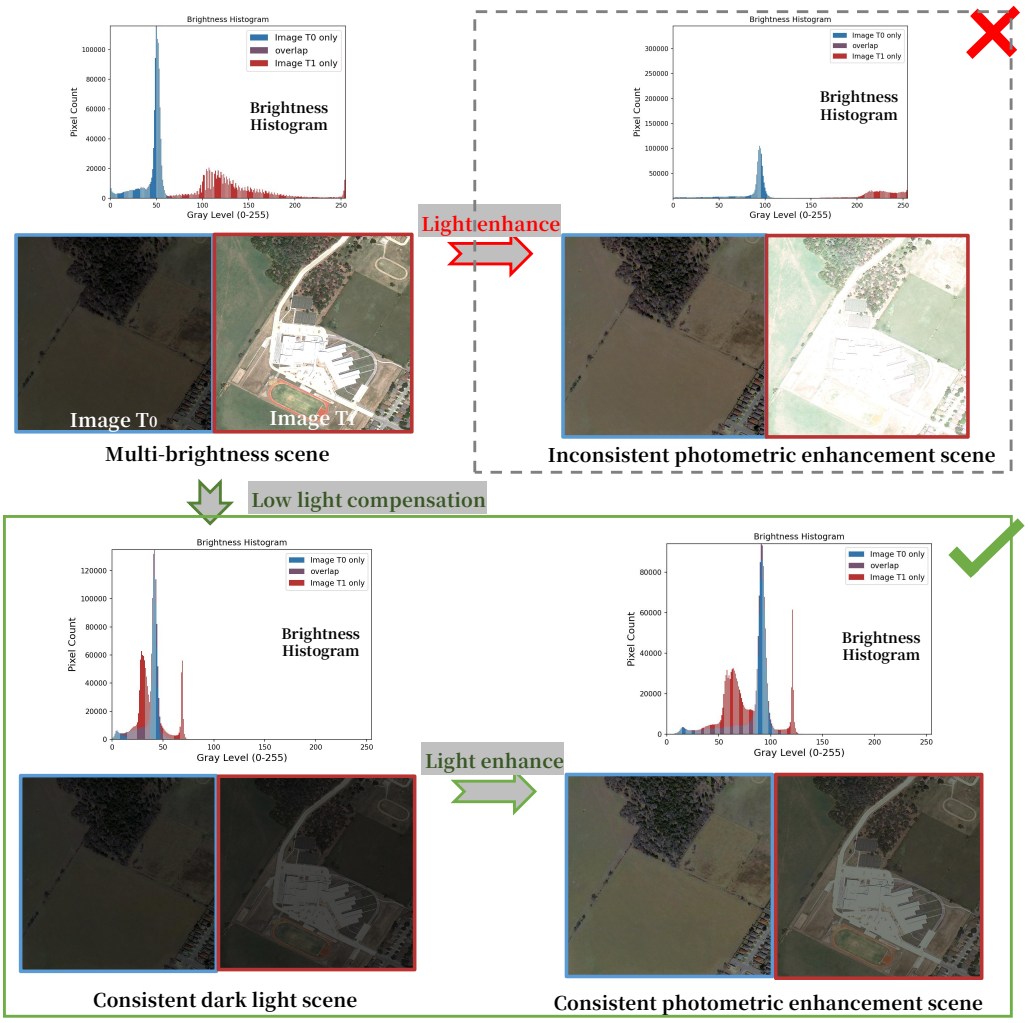


Figure 12. Comparison between independent photometric enhancement and the proposed compensation-based strategy on the LEVIR-CD dataset. This figure highlights that applying a generic enhancement model to images with different brightness levels may over-correct normal-light regions and suppress useful high-intensity details, while the compensation-based strategy better preserves cross-temporal photometric consistency.

4.3.2. Study on Bi-Temporal Enhanced Module

This section primarily investigates the functionality and importance of the bi-temporal enhanced module through analysis on three representative data scenarios. As shown in Figure 13, the first, third, and fifth rows present change detection results obtained by directly feeding the consistent dark-light scenes generated by the low-light compensation mechanism module into the change detection network. In contrast, the second, fourth, and sixth rows show the corresponding results obtained by applying the bi-temporal enhanced module to generate enhanced images prior to change detection.

The first and second columns display the bi-temporal remote sensing images, while the third, fourth, and fifth columns visualize the corresponding attention maps produced by the change detection network. The sixth column presents the ground-truth reference maps, and the seventh column shows the final change detection results. For clearer visualization, different colors are used to indicate classification outcomes, where white denotes true positives, black denotes true negatives, red denotes false positives, and green denotes false negatives.

As observed from the results, both the consistent dark-light images and the enhanced images enable the network to roughly localize the changed regions, indicating that the

framework possesses a certain capability to identify change-prone areas even under low-light conditions. However, after applying the bi-temporal enhanced module, the feature extraction of fine-grained texture details is significantly improved. In particular, the boundaries and structural details of changed regions are delineated more accurately, leading to fewer false detections along object contours. These observations demonstrate that the bi-temporal enhanced module effectively enhances photometric and structural representations, thereby providing higher-quality inputs for subsequent change detection and improving overall detection accuracy under complex illumination conditions.

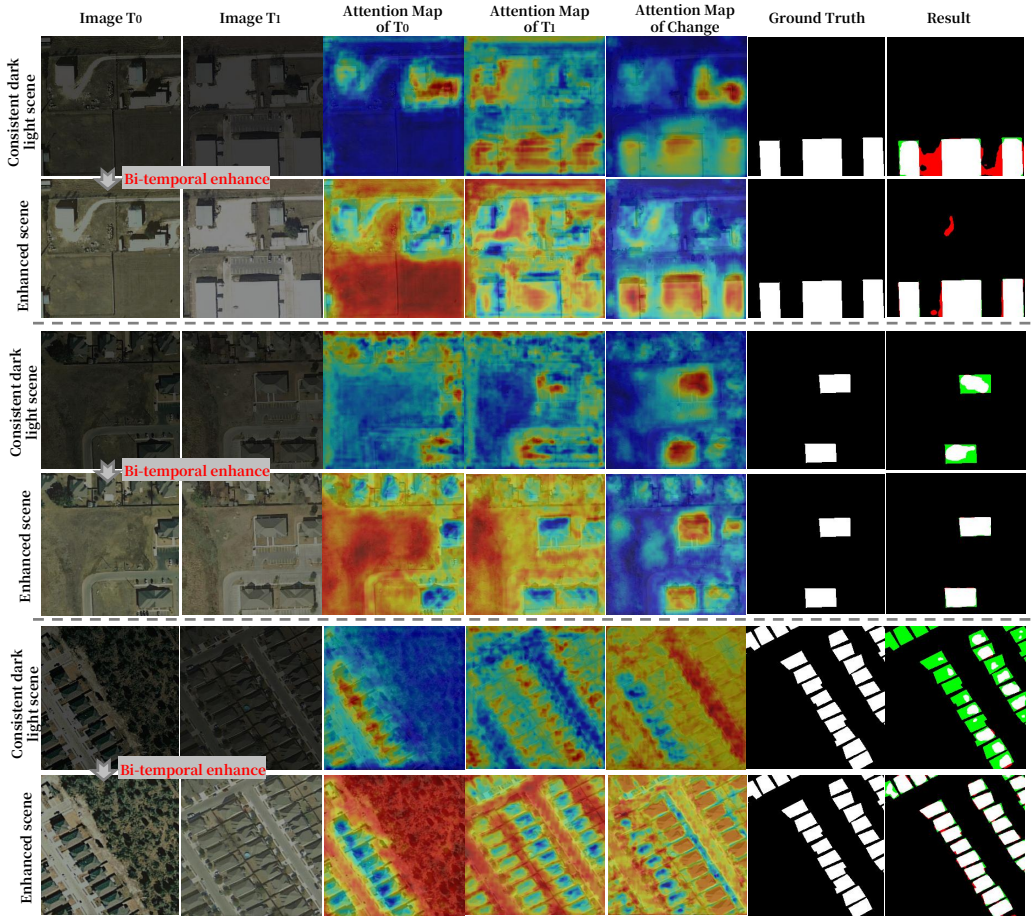


Figure 13. Visualization of the effectiveness of the bi-temporal enhanced module. Change detection results obtained from consistent dark-light scenes generated by the low-light compensation mechanism module and the enhanced scenes produced by the bi-temporal enhanced module. In the results, white denotes true positives, black denotes true negatives, red denotes false positives, and green denotes false negatives.

4.3.3. Loss Function

Table 3 presents the overall performance of the model under different loss function combinations. By first comparing the first and second rows, as well as the fourth and fifth rows, it can be observed that the loss associated with the LLCMM module plays a crucial role in improving overall detection accuracy. This improvement mainly arises because the loss function performs dark-light compensation on image pairs with large brightness discrepancies, thereby preventing overexposure when bright images are directly subjected to photometric enhancement. As a result, the overall stability of the model is enhanced.

Further comparison between the first and fourth rows indicates that the BTEM with GT-Light enhancement loss yields a pronounced performance gain. The photometric en-

hancement effectively addresses the weak texture representation under low-light conditions, enabling more sufficient and effective feature learning from the enhanced textures.

Finally, comparisons between the second and third rows, as well as between the fifth and sixth rows, show that the BTEM with Light Consistent Loss provides only marginal improvements in overall accuracy. This suggests that after dark-light compensation or photometric enhancement, the additional photometric consistency constraint contributes limited improvement to image quality. Overall, these results demonstrate that the designed loss functions play an important role during model training and jointly contribute to the performance gains of the proposed framework.

4.3.4. Different Gamma Correction Analysis

To evaluate the robustness of the proposed framework under different low-light conditions and to analyze the contribution of each module, we conduct experiments with different gamma values. The quantitative results in Table 4 indicate that, under all tested illumination levels, using only the Change Detection Module and the Bi-temporal Enhanced Module is insufficient to achieve stable and satisfactory detection performance. As further illustrated in Figure 14, image enhancement alone, without explicitly aligning the brightness of the bi-temporal images, cannot effectively eliminate cross-temporal illumination inconsistency. In some cases, applying enhancement to images that already have normal brightness may even lead to overexposure, which enlarges the feature discrepancy of similar land-cover objects and negatively affects the final detection results. By contrast, the proposed framework remains more stable across different gamma settings.

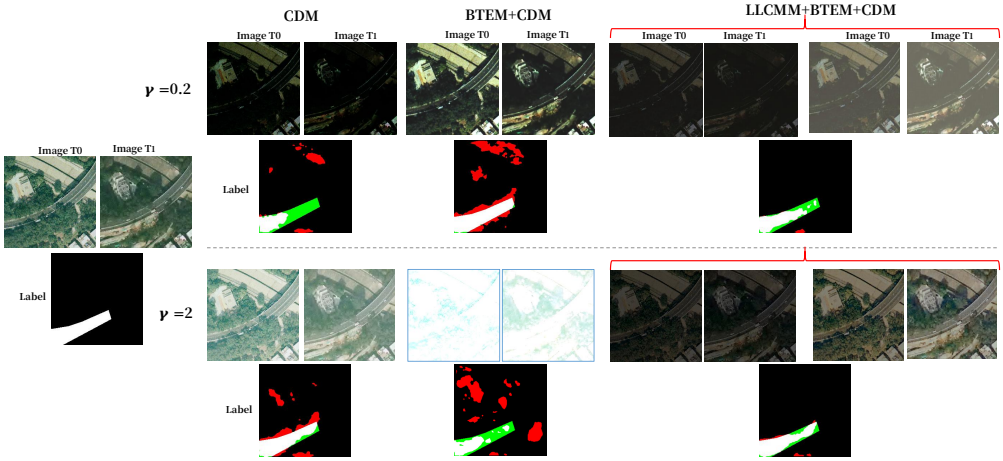


Figure 14. Qualitative visualization demonstrating the effectiveness of individual modules on representative samples from the SYSU dataset under different gamma settings. In the results, white denotes true positives, black denotes true negatives, red denotes false positives, and green denotes false negatives. Note: LLCMM = Low-light Compensation Mechanism Module; BTEM = Bi-temporal Enhanced Module; CDM = Change Detection Module.

Table 4. Quantitative ablation study results demonstrating the effectiveness of individual modules on representative samples from the SYSU dataset under different gamma settings.

Modules (✓ = Enabled)			gamma = 0.2		gamma = 2	
LLCMM	**BTEM**	**CDM**	**F1 ↑**	**IoU ↑**	**F1 ↑**	**IoU ↑**
		✓	51.08	45.71	65.89	56.42
	✓	✓	68.71	61.56	50.64	43.91
✓	✓	✓	**76.42**	**63.41**	**78.15**	**65.33**

Note: LLCMM = Low-light Compensation Mechanism Module; BTEM = Bi-temporal Enhanced Module; CDM = Change Detection Module. ✓ indicates that the corresponding module is enabled and the best results are highlighted in **bold** . ↑ indicates that higher values are better.

4.3.5. Efficiency and Parameter Sensitivity Analysis

Efficiency Analysis. To further evaluate the computational efficiency of CPEN, we compare it with representative baselines in terms of the number of parameters, FLOPs, and FPS, as reported in Table 2. This analysis provides a more comprehensive assessment of the proposed framework from the perspective of computational cost. The results show that, among the direct change detection baselines, CPEN has the smallest number of parameters. Although its FLOPs are slightly higher and the corresponding FPS is slightly lower, it still maintains relatively fast inference in practice. More importantly, compared with the enhancement-assisted baselines, CPEN shows clear advantages in parameter count, computational complexity, and FPS. This indicates that the proposed unified framework achieves a better trade-off between detection performance and efficiency than cascaded "enhancement + detection" pipelines. Furthermore, the module-wise analysis in Table 3 shows that most of the parameters are concentrated in the Change Detection Module, while the Low-light Compensation Mechanism Module and the Bi-temporal Enhanced Module introduce only marginal parameter increases and almost no additional impact on FPS.

This demonstrates that the two illumination-oriented modules improve robustness under low-light conditions with very limited extra computational cost, while the overall efficiency of CPEN is still mainly determined by the change detection backbone.

Since the proposed framework adopts a stage-wise optimization strategy and different hyperparameters are involved in different training stages, we conduct a focused sensitivity analysis on the key hyperparameters in the first-stage training schedule, including the loss weights and the stage transition epochs. Specifically, we vary only one parameter at a time while keeping the others fixed to the default setting. As shown in Table 5, the default setting achieves 90.39% F1 and 82.67% IoU on LEVIR-CD, while the tested variants obtain F1 scores from 89.64% to 90.88% and IoU values from 81.76% to 82.91%. The performance remains consistently strong across all settings, indicating that our model is not highly sensitive to these hyperparameters and that the proposed stage-wise training strategy is relatively robust.

Table 5. Sensitivity analysis of stage-wise training hyperparameters on the LEVIR-CD dataset. Only one parameter is varied at a time while others are fixed to the default setting ($\lambda_1 = 0.6$, $\lambda_2 = 0.3$, $\lambda_3 = 0.1$, $N^1_{epoch} = 1000$, $N^2_{epoch} = 10{,}000$). ↑ indicates that higher values are better.

Setting	F1 ↑	IoU ↑
Baseline	90.39	82.67
$\lambda_1 = 0.4$	89.64	81.76
$\lambda_1 = 0.8$	90.68	82.83
$\lambda_2 = 0.2$	90.52	82.90
$\lambda_2 = 0.4$	89.96	81.94
$\lambda_3 = 0.05$	90.17	82.45
$\lambda_3 = 0.2$	89.84	81.89
$N^1_{epoch} = 800$	90.73	82.46
$N^1_{epoch} = 1200$	90.88	82.37
$N^2_{epoch} = 8000$	90.52	82.91
$N^2_{epoch} = 12{,}000$	90.12	82.04

5. Discussion

Simulated and Realistic Dataset Discussion: Although the proposed synthetic low-light generation strategy provides a controllable way to simulate brightness degradation, it

still cannot fully reflect the complex degradations in real low-light remote sensing scenarios. In particular, real data may involve spatially varying illumination, shadow interference, geometry-related occlusions, low-light noise, and imaging pipeline distortions, which are not fully represented in the current synthetic setting. When evaluated on the real-world NPULL-CD dataset, a model trained on PRCV achieves 73.5% in F1, suggesting a noticeable domain gap between synthetic and real low-light data. This limitation indicates that synthetic data are better viewed as a controllable complement to rather than a replacement for the real-world data. This also explains why CPEN obtains higher scores on the simulated LEVIR-CD dataset (90.39% F1 and 82.67% IoU) than on the real-world NPULL-CD dataset (86.26% F1 and 79.02% IoU). LEVIR-CD is degraded by a controlled gamma transform, while its original geometry, object boundaries, and annotations are inherited from a clean benchmark. Therefore, the main difficulty in this setting is a known radiometric perturbation. NPULL-CD, by contrast, is collected under real daytime–nightfall conditions, where nonuniform illumination, shadows, low-light noise, exposure and imaging-pipeline differences, and residual registration errors are coupled together. These factors can introduce ambiguous boundary pixels and false responses in unchanged regions. Thus, the lower score on NPULL-CD reflects the additional difficulty of real low-light acquisition rather than a contradiction in the evaluation. The gamma-based datasets are used for controlled robustness and generalization tests, while NPULL-CD is used to evaluate practical applicability under real illumination inconsistency. Future work will focus on collecting more diverse real low-light datasets and developing more realistic degradation modeling for improved generalization.

Another important future direction is to analyze illumination sensitivity across different land-cover types, such as buildings, vegetation, roads, and water bodies. Different materials and surface structures may respond differently to low-light acquisition, shadows, and exposure variation. However, the datasets used in this work mainly provide binary change masks rather than reliable semantic land-cover annotations for each pixel. Therefore, we do not report class-wise illumination sensitivity statistics in the current revision to avoid unsupported or misleading conclusions. In future work, we will combine low-light change detection with semantic land-cover labels or external land-cover masks to evaluate class-wise robustness under illumination variation more systematically.

6. Conclusions

This paper presented a Consistent Photometric Enhancement Network (CPEN) for remote sensing change detection under complex illumination conditions. By jointly performing low-light compensation, bi-temporal enhancement, and change detection, the proposed framework reduces photometric inconsistency between image pairs and improves the reliability of feature representation. Experimental results on the multi-brightness benchmark show that CPEN achieves good performance compared with existing methods, especially under severe illumination variation, while ablation studies confirm the contribution of each module. In future work, we will collect more real low-light remote sensing data to further evaluate the proposed method and improve its generalization in real-world scenarios.

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

Funding: This research was funded in part by the Postdoctoral Fellowship Program of CPSF, grant number GZB20240986; the Shaanxi Postdoctoral Research Project Funding Program, grant number W016375, and the China Scholarship Council (CSC), grant number 202506290204.

Data Availability Statement: The public datasets analyzed in this study are available from their original providers. The NPULL-CD dataset and implementation code are available from the first author upon reasonable request.

Conflicts of Interest: The authors declare no conflicts of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript; or in the decision to publish the results.

Abbreviations

The following abbreviations are used in this manuscript:

BIT	Bi-Temporal Interaction Transformer
BTEM	Bi-Temporal Enhanced Module
CDM	Change Detection Module
CPEN	Consistent Photometric Enhancement Network
GCAM	Global Cross-Attention Module
IoU	Intersection over Union
LLCMM	Low-Light Compensation Mechanism Module
LLIE	Low-Light Image Enhancement
NPULL-CD	Northwestern Polytechnical University Low-Light Change Dataset
PEM	Pixel-wise Enhancement Module

References

1. Huang, X.; Zhang, L.; Zhu, T. Building Change Detection From Multitemporal High-Resolution Remotely Sensed Images Based on a Morphological Building Index. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2014**, *7*, 105. [CrossRef]
2. Rokni, K.; Ahmad, A.; Selamat, A.; Hazini, S. Water feature extraction and change detection using multitemporal Landsat imagery. *Remote Sens.* **2014**, *6*, 4173–4189. [CrossRef]
3. Gong, M.; Zhao, J.; Liu, J.; Miao, Q.; Jiao, L. Change detection in synthetic aperture radar images based on deep neural networks. *IEEE Trans. Neural Netw. Learn. Syst.* **2015**, *27*, 125–138. [PubMed]
4. Chen, C.F.; Son, N.T.; Chang, N.B.; Chen, C.R.; Chang, L.Y.; Valdez, M.; Centeno, G.; Thompson, C.A.; Aceituno, J.L. Multi-decadal mangrove forest change detection and prediction in Honduras, Central America, with Landsat imagery and a Markov chain model. *Remote Sens.* **2013**, *5*, 6408–6426. [CrossRef]
5. Jiang, H.; Peng, M.; Zhong, Y.; Xie, H.; Hao, Z.; Lin, J.; Ma, X.; Hu, X. A survey on deep learning-based change detection from high-resolution remote sensing images. *Remote Sens.* **2022**, *14*, 1552. [CrossRef]
6. Daudt, R.C.; Le Saux, B.; Boulch, A. Fully convolutional siamese networks for change detection. In *Proceedings of the 2018 25th IEEE International Conference on Image Processing (ICIP)*; IEEE: Piscataway, NJ, USA, 2018; pp. 4063–4067.
7. Chen, H.; Qi, Z.; Shi, Z. Remote sensing image change detection with transformers. *IEEE Trans. Geosci. Remote Sens.* **2021**, *60*, 5607514.
8. Fang, S.; Li, K.; Shao, J.; Li, Z. SNUNet-CD: A Densely Connected Siamese Network for Change Detection of VHR Images. *IEEE Geosci. Remote Sens. Lett.* **2022**, *19*, 3056416. [CrossRef]
9. Zhang, H.; Chen, K.; Liu, C.; Chen, H.; Zou, Z.; Shi, Z. CDMamba: Incorporating local clues into mamba for remote sensing image binary change detection. *IEEE Trans. Geosci. Remote Sens.* **2025**, *63*, 4405016. [CrossRef]
10. Fu, Y.; Hong, Y.; Chen, L.; You, S. LE-GAN: Unsupervised low-light image enhancement network using attention module and identity invariant loss. *Knowl.-Based Syst.* **2022**, *240*, 108010. [CrossRef]
11. Lore, K.G.; Akintayo, A.; Sarkar, S. LLNet: A deep autoencoder approach to natural low-light image enhancement. *Pattern Recognit.* **2017**, *61*, 650–662. [CrossRef]
12. Zamir, S.W.; Arora, A.; Khan, S.; Hayat, M.; Khan, F.S.; Yang, M.H.; Shao, L. Learning enriched features for real image restoration and enhancement. In *Proceedings of the Computer Vision–ECCV 2020: 16th European Conference, Glasgow, UK, 23–28 August 2020*; Springer: Berlin/Heidelberg, Germany, 2020; Part XXV 16, pp. 492–511.
13. Zhou, S.; Li, C.; Change Loy, C. Lednet: Joint low-light enhancement and deblurring in the dark. In *Proceedings of the European Conference on Computer Vision*; Springer: Berlin/Heidelberg, Germany, 2022; pp. 573–589.

14. Cai, Y.; Bian, H.; Lin, J.; Wang, H.; Timofte, R.; Zhang, Y. Retinexformer: One-stage retinex-based transformer for low-light image enhancement. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*; IEEE: Piscataway, NJ, USA, 2023; pp. 12504–12513.
15. Rahaman, N.; Baratin, A.; Arpit, D.; Draxler, F.; Lin, M.; Hamprecht, F.; Bengio, Y.; Courville, A. On the spectral bias of neural networks. In Proceedings of the International Conference on Machine Learning, PMLR, Long Beach, CA, USA, 9–15 June 2019 ; pp. 5301–5310.
16. Wu, C.; Zhang, L.; Zhang, L. A scene change detection framework for multi-temporal very high resolution remote sensing images. *Signal Process.* **2016**, *124*, 184–197. [CrossRef]
17. Wang, Y.; Du, B.; Ru, L.; Wu, C.; Luo, H. Scene change detection via deep convolution canonical correlation analysis neural network. In *Proceedings of the IGARSS 2019—2019 IEEE International Geoscience and Remote Sensing Symposium*; IEEE: Piscataway, NJ, USA, 2019; pp. 198–201.
18. Wu, C.; Zhang, L.; Du, B. Kernel slow feature analysis for scene change detection. *IEEE Trans. Geosci. Remote Sens.* **2017**, *55*, 2367–2384. [CrossRef]
19. Wang, Q.; Zhang, X.; Chen, G.; Dai, F.; Gong, Y.; Zhu, K. Change detection based on Faster R-CNN for high-resolution remote sensing images. *Remote Sens. Lett.* **2018**, *9*, 923–932. [CrossRef]
20. Zhang, L.; Hu, X.; Zhang, M.; Shu, Z.; Zhou, H. Object-level change detection with a dual correlation attention-guided detector. *ISPRS J. Photogramm. Remote Sens.* **2021**, *177*, 147–160. [CrossRef]
21. Han, P.; Ma, C.; Li, Q.; Leng, P.; Bu, S.; Li, K. Aerial image change detection using dual regions of interest networks. *Neurocomputing* **2019**, *349*, 190–201. [CrossRef]
22. Hou, B.; Liu, Q.; Wang, H.; Wang, Y. From W-Net to CDGAN: Bitemporal Change Detection via Deep Learning Techniques. *IEEE Trans. Geosci. Remote Sens.* **2020**, *58*, 1790–1802. [CrossRef]
23. Chen, H.; Shi, Z. A spatial-temporal attention-based method and a new dataset for remote sensing image change detection. *Remote Sens.* **2020**, *12*, 1662. [CrossRef]
24. Li, S.; Han, P.; Bu, S.; Tong, P.; Li, Q.; Li, K.; Wan, G. Change detection in images using shape-aware siamese convolutional network. *Eng. Appl. Artif. Intell.* **2020**, *94*, 103819. [CrossRef]
25. Feng, Y.; Jiang, J.; Xu, H.; Zheng, J. Change detection on remote sensing images using dual-branch multilevel intertemporal network. *IEEE Trans. Geosci. Remote Sens.* **2023**, *61*, 4401015. [CrossRef]
26. Huang, Y.; Li, X.; Du, Z.; Shen, H. Spatiotemporal enhancement and interlevel fusion network for remote sensing images change detection. *IEEE Trans. Geosci. Remote Sens.* **2024**, *62*, 5609414. [CrossRef]
27. Feng, Y.; Qin, M.; Jiang, J.; Lai, J.; Zheng, J. Axial-shunted spatial-temporal conversation for change detection. *ACM Trans. Multimed. Comput. Commun. Appl.* **2025**, *21*, 1–21. [CrossRef]
28. Weng, L.; Yang, W.; Hu, B.; Han, P.; Xue, S.; Zhang, Y.; Li, H.; Jin, J.; Bu, S. MDINet: Multi-Domain Incremental Network for Change Detection. *IEEE Trans. Geosci. Remote Sens.* **2024**, *62*, 4402315. [CrossRef]
29. Liu, Y.; He, Q.; Li, J.; Liu, X.; Fiorio, P.R.; Nakai, É.S.; Yang, B. Towards resolution-arbitrary remote sensing change detection with Spatial-frequency dual domain learning. *ISPRS J. Photogramm. Remote Sens.* **2026**, *231*, 137–150. [CrossRef]
30. Chen, H.; Lan, C.; Song, J.; Ibañez, D.; Xia, J.; Schindler, K.; Yokoya, N. Multimodal remote sensing change detection: An image matching perspective. *ISPRS J. Photogramm. Remote Sens.* **2026**, *233*, 487–501. [CrossRef]
31. Celik, T.; Tjahjadi, T. Contextual and variational contrast enhancement. *IEEE Trans. Image Process.* **2011**, *20*, 3431–3441. [CrossRef] [PubMed]
32. Rahman, S.; Rahman, M.M.; Abdullah-Al-Wadud, M.; Al-Quaderi, G.D.; Shoyaib, M. An adaptive gamma correction for image enhancement. *EURASIP J. Image Video Process.* **2016**, *2016*, 35. [CrossRef]
33. Jobson, D.J.; Rahman, Z.; Woodell, G.A. Properties and performance of a center/surround retinex. *IEEE Trans. Image Process.* **1997**, *6*, 451–462. [CrossRef] [PubMed]
34. Li, M.; Liu, J.; Yang, W.; Sun, X.; Guo, Z. Structure-revealing low-light image enhancement via robust retinex model. *IEEE Trans. Image Process.* **2018**, *27*, 2828–2841. [CrossRef] [PubMed]
35. Guo, X.; Li, Y.; Ling, H. LIME: Low-light image enhancement via illumination map estimation. *IEEE Trans. Image Process.* **2016**, *26*, 982–993. [CrossRef] [PubMed]
36. Wei, C.; Wang, W.; Yang, W.; Liu, J. Deep retinex decomposition for low-light enhancement. *arXiv* **2018**, arXiv:1808.04560.
37. Zhang, Y.; Zhang, J.; Guo, X. Kindling the darkness: A practical low-light image enhancer. In Proceedings of the 27th ACM International Conference on Multimedia, Nice, France, 21–25 October 2019; pp. 1632–1640.
38. Wang, R.; Zhang, Q.; Fu, C.W.; Shen, X.; Zheng, W.S.; Jia, J. Underexposed photo enhancement using deep illumination estimation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*; IEEE: Piscataway, NJ, USA, 2019; pp. 6849–6857.
39. Lei, L.; Sun, Y.; Kuang, G. Adaptive Local Structure Consistency-Based Heterogeneous Remote Sensing Change Detection. *IEEE Geosci. Remote Sens. Lett.* **2022**, *19*, 8003905. [CrossRef]

40. Sun, Y.; Lei, L.; Guan, D.; Kuang, G. Iterative Robust Graph for Unsupervised Change Detection of Heterogeneous Remote Sensing Images. *IEEE Trans. Image Process.* **2021**, *30*, 6277–6291. [CrossRef] [PubMed]
41. Sun, Y.; Lei, L.; Zhang, L.; Kuang, K.; Yu, G. Detecting Changes Without Comparing Images: Rules Induced Change Detection in Heterogeneous Remote Sensing Images. *ISPRS J. Photogramm. Remote Sens.* **2025**, *230*, 241–257. [CrossRef]
42. Goodfellow, I.; Bengio, Y.; Courville, A. *Deep Learning*; MIT Press: Cambridge, MA, USA, 2016.
43. Yang, S.; Ding, M.; Wu, Y.; Li, Z.; Zhang, J. Implicit neural representation for cooperative low-light image enhancement. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*; IEEE: Piscataway, NJ, USA, 2023; pp. 12918–12927.
44. Ba, J.L.; Kiros, J.R.; Hinton, G.E. Layer normalization. *arXiv* **2016**, arXiv:1607.06450.
45. Touvron, H.; Cord, M.; Sablayrolles, A.; Synnaeve, G.; Jégou, H. Going deeper with image transformers. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*; IEEE: Piscataway, NJ, USA, 2021; pp. 32–42.
46. Hendrycks, D.; Gimpel, K. Gaussian error linear units (gelus). *arXiv* **2016**, arXiv:1606.08415.
47. Yang, K.; Xia, G.S.; Liu, Z.; Du, B.; Yang, W.; Pelillo, M.; Zhang, L. Semantic change detection with asymmetric Siamese networks. *arXiv* **2020**, arXiv:2010.05687.
48. Shi, Q.; Liu, M.; Li, S.; Liu, X.; Wang, F.; Zhang, L. A deeply supervised attention metric-based network and an open aerial image dataset for remote sensing change detection. *IEEE Trans. Geosci. Remote Sens.* **2021**, *60*, 5604816. [CrossRef]
49. Yan, Q.; Feng, Y.; Zhang, C.; Pang, G.; Shi, K.; Wu, P.; Dong, W.; Sun, J.; Zhang, Y. Hvi: A new color space for low-light image enhancement. In Proceedings of the Computer Vision and Pattern Recognition Conference, Nashville, TN, USA, 11–15 June 2025; pp. 5678–5687.

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.