electronics

MDPI

Article

A Light-Weight Self-Supervised Infrared Image Perception Enhancement Method

Yifan Xiao [1,2], Zhilong Zhang [1,*] and Zhouli Li [2]

[1] National Laboratory of Automatic Target Recognition, National University of Defense Technology, Changsha 410073, China; xiaoyifan.xyf@gmail.com
[2] College of Chemistry and Chemical Engineering, Xi'an Shiyou University, Xi'an 710065, China; lizhouli@xsyu.edu.cn
* Correspondence: zhangzhilong@nudt.edu.cn

Abstract: Convolutional Neural Networks (*CNNs*) have achieved remarkable results in the field of infrared image enhancement. However, the research on the visual perception mechanism and the objective evaluation indicators for enhanced infrared images is still not in-depth enough. To make the subjective and objective evaluation more consistent, this paper uses a perceptual metric to evaluate the enhancement effect of infrared images. The perceptual metric mimics the early conversion process of the human visual system and uses the normalized Laplacian pyramid distance (*NLPD*) between the enhanced image and the original scene radiance to evaluate the image enhancement effect. Based on this, this paper designs an infrared image-enhancement algorithm that is more conducive to human visual perception. The algorithm uses a lightweight Fully Convolutional Network (*FCN*), with *NLPD* as the similarity measure, and trains the network in a self-supervised manner by minimizing the *NLPD* between the enhanced image and the original scene radiance to achieve infrared image enhancement. The experimental results show that the infrared image enhancement method in this paper outperforms existing methods in terms of visual perception quality, and due to the use of a lightweight network, it is also the fastest enhancement method currently.

Keywords: visual perception; normalized Laplacian pyramid; self-supervision; lightweight

Citation: Xiao, Y.; Zhang, Z.; Li, Z. A Light-Weight Self-Supervised Infrared Image Perception Enhancement Method. *Electronics* **2024**, *13*, 3695. https://doi.org/10.3390/electronics13183695

Academic Editors: Silvia Liberata Ullo and Li Zhang

Received: 23 August 2024
Revised: 6 September 2024
Accepted: 8 September 2024
Published: 18 September 2024

1. Introduction

The realm of infrared image capture has attracted considerable interest in recent years, primarily fueled by its widespread applications in defense, surveillance, and various other sectors. Nonetheless, infrared images frequently encounter challenges such as low contrast and blurred details. Specifically, the low resolution of infrared images and the existence of low clouds can obstruct the detection of heat-emitting objects, limiting the efficacy of observing infrared targets and impeding the progress of infrared imaging applications. Therefore, achieving high-quality infrared images necessitates enhancement [1,2]. Enhancement techniques aim to improve image quality, enhance clarity, and boost visual impact, playing a crucial role in tasks such as object recognition [3], instance segmentation [4], tracking, and detection [5,6], among others. The existing methods for enhancing infrared images can be broadly classified into two groups: traditional methodologies and deep learning approaches.

Throughout the past few decades, classical methods such as histogram equalization [7–9] and gamma correction [10] have demonstrated efficacy in enhancing low-light images [11,12]. Moreover, a variety of classical techniques rooted in the Retinex theory [13] have been devised, integrating various prior regularization optimization models to separate the illumination and reflectance image layers' structures [14–16]. However, these manually crafted constraints and priors lack adaptability, potentially resulting in outcomes marked by noticeable noise or influenced by excessive or insufficient enhancement, which could adversely impact human visual perception.

In recent years, propelled by rapid advancements in deep learning, an increasing number of researchers have harnessed this technology in the field of image enhancement [17]. These methods, based on distinct learning approaches, can be classified into supervised learning [18–20] and unsupervised learning [21,22]. However, the effectiveness of these deep learning techniques heavily relies on intricately designed architectures and meticulously curated training datasets. Consequently, while excelling in objective metrics, they often struggle to align subjective visual experiences with objective assessments in practical scenarios.

The essence of crafting a successful enhancement algorithm, whether through classical methodologies or deep learning approaches, lies in the capacity to extract and safeguard relevant information within the image, while simultaneously eliminating redundancy and noise, all while conforming to human perceptual standards. This foundational principle forms the core of our design philosophy. Historically, discrepancies have frequently arisen in the evaluation metrics utilized for image enhancement between subjective and objective assessments. While many techniques excel in objective evaluations, a discernible gap often persists when assessed against human perception subjectively. Therefore, bridging this disparity between subjective and objective evaluations necessitates the implementation of an efficient methodology for enhancing infrared perception.

With the objective of enhancing infrared perception, we have introduced a self-supervised algorithm. Departing from traditional supervised learning methods, our approach moves away from relying on actual data for supervised training. Instead, we utilize the estimation of the original scene's radiation intensity range from the input image as the training ground truth. Moreover, we have developed a no-reference quality assessment metric based on perceptual criteria. This metric utilizes the $NLPD$ between the image and the original scene's radiation intensity to evaluate the perceptual enhancement impact of the image [23], simulating the initial transformations within the human visual system [24]. Expanding on this framework, we incorporated $NLPD$ into a CNN to enhance the perceptual similarity between the image and the original scene's radiation intensity by minimizing the perceptual loss metric. The CNN we implemented boasts a lightweight structure capable of preserving intricate image details while enabling swift online deployment. This design achieves commendable performance at a minimal computational cost.

Our contributions can be succinctly summarized as follows:

- Owing to the specific heat distribution and object features contained in infrared images, a lightweight FCN structure is designed to capture the key information in infrared images, such as hotspots, edges, and textures.
- By training the model in a self-supervised manner, the proposed method overcomes the limitation of traditional supervised learning, which requires a large amount of ground truth data. This reduces the data cost and workload and improves the utilization rate of the available data.
- By incorporating $NLPD$ into the loss function of the CNN, the proposed method leverages the multi-scale image details extracted by the normalized Laplacian pyramid. This enables the enhancement model to achieve excellent results in infrared image perceptual enhancement and demonstrates robust and generalized performance.
- Our method achieves excellent performance with a small computational cost and has the fastest running speed, making it suitable for a wider range of physical scenarios.

Through these advancements, our infrared image enhancement technique surpasses existing methods in both visual perception quality and operational speed, promising broad application prospects in various domains.

The rest of this paper is organized as follows: Section 2 reviews related works on traditional image enhancement and perceptual image enhancement. Section 3 details the proposed approach and its underlying principles. Section 4 describes the experimental results and analysis. Section 5 discusses the experimental findings. Finally, Section 6 summarizes the research insights and provides an outlook on future work.

2. Related Work

2.1. Traditional Image Enhancement Methods

Numerous conventional enhancement algorithms have been developed to produce high-contrast infrared images. Among these techniques, histogram equalization (HE) [25] emerges as a foundational method that redistributes the grayscale values of image pixels to achieve a more uniform distribution across the entire grayscale spectrum. This technique enhances the contrast and brightness distribution of the image, proving especially effective in visible light imagery. Nonetheless, in the context of noisy infrared images, HE often amplifies both the desired signal and the noise simultaneously, leading to suboptimal results. To mitigate this challenge, two variants of HE have been introduced:

Dynamic Histogram Equalization and Contrast Enhancement ($DHECI$) [26]: $DHECI$ integrates dynamic histogram equalization with contrast enhancement to improve both contrast and detail information in images. This technique seeks to enhance the overall effect by incorporating dynamic adjustments and contrast enhancements, proving especially effective in preserving details and enhancing contrast, particularly in hyperspectral images. Contrast-Limited Adaptive Histogram Equalization ($CLAHE$) [27]: $CLAHE$ is an adaptive HE technique that divides the image into multiple local regions and applies the HE algorithm to each region separately, avoiding potential issues of over-enhancement that global enhancement methods might face. This method performs exceptionally well in situations where images display uneven brightness distributions.

Moreover, Natural Picture Enhancement (NPE) [28] is designed to enhance the visual quality of natural scene images by refining contrast and brightness levels. NPE enriches image details and colors, enhancing visual appeal and clarity. Unlike $CLAHE$, NPE is a versatile enhancement technique suitable for a wide range of natural scene images. By manipulating parameters like $gamma$ [29] (e.g., $gamma = 0.8$), users can customize contrast and brightness adjustments to achieve desired image enhancement effects. However, these methods may inadvertently introduce increased detail blurring while suppressing noise, potentially compromising the overall quality of the image.

Additionally, classic methods based on wavelet transform [30] and low-light image enhancement ($LIME$) [31] have been developed:

Wavelet Transform: This approach involves decomposing input images with infrared noise into multiple scales, denoising each scale by utilizing a soft threshold, and subsequently converting the processed wavelet coefficients back to the spatial domain. Low-Light Image Enhancement ($LIME$): $LIME$ is specifically designed to improve images captured under low-light conditions. It focuses on enhancing brightness and clarity to enhance visibility in dim environments. $LIME$ effectively boosts the quality of low-light images by adjusting relevant parameters, thereby enhancing overall visibility and image quality. In contrast to wavelet-based denoising techniques, $LIME$ emphasizes enhancing visual effects rather than solely processing noise; while these methods are effective in preventing noise amplification, the extent of their enhancement impact may vary.

In recent years, propelled by the rapid progress of deep learning, a growing cohort of researchers has begun integrating it into the realm of image enhancement. Supervised learning approaches commonly leverage paired images to grasp the mapping relationship from suboptimal illumination images to those captured under normal illumination conditions. The MLLEN-IC network framework, as introduced by Fan et al. [32], incorporates the core unit of the Rse2Net constructed network to boost model efficacy through the extraction of multi-scale features. Yang et al. introduced the Deep Recursive Band Network ($DRBN$) for reconstructing improved normal illumination images from paired low-light images utilizing a linear band representation [33]. Furthermore, the URetinex-Net network framework, proposed by Wu et al., decomposes low-light images into reflection layers and illumination layers to amplify illumination effects [34]. However, training these techniques on paired datasets can potentially result in overfitting, limiting the model's ability to generalize effectively. In a departure from the dependence on paired data, Jiang et al. introduced an unsupervised Generative Adversarial Network ($EnlightenGAN$) that operates without

the need for paired low-light or normal-light image training. Instead, it regularizes based on information gleaned from inputs [35]. This approach of unpaired training regularization utilizing input-extracted information adeptly tackles various novel challenges in low-light image enhancement. Likewise, Liu et al. developed intrinsic exposure structures that delineate weak light images grounded on Retinex principles. They devised a lightweight and efficient low-light image enhancement network by uncovering low-light prior structures from a condensed search space through collaborative learning without reference [36]. Li et al. introduced Zero-Reference Deep Curve Estimation ($Zero-DceP$), a method that enables training without paired data by employing a non-reference loss function, effectively addressing cutting-edge challenges in low-light image enhancement [37]. Nevertheless, the efficacy of these deep learning techniques is heavily contingent on intricately crafted architectures and thoughtfully chosen training data, frequently leading to subpar generalization in real-world settings.

2.2. Image Perception Enhancement Methods

When assessing image quality, metrics like $PSNR$ [38], $SSIM$ [39], FMI [40], and others frequently exhibit notable disparities in comparison to subjective evaluations, underscoring the intricacies of visual perceptual processes. Minimizing or even eradicating this gap poses a fundamental challenge for contemporary image enhancement methods.

Laparra et al. introduced perceptual distance, utilizing the $NLPD$ to simulate the initial processing stages of the human visual system. By minimizing the perceived differences between generated images and the original scenes, image enhancement is accomplished, framing the task as a constrained optimization problem. However, due to the non-convex nature of $NLPD$ and the high dimensionality of constrained optimization problems, the gradient-based iterative solving methods initially proposed frequently present formidable computational hurdles; while its image enhancement strategy aligns with human visual perception, the slow processing speed hampers its practical applicability in various scenarios.

$SRCNN$ [41] was one of the pioneering works to employ CNN for image super-resolution enhancement, leveraging the benefits of full convolution and a lightweight design. Nonetheless, relying solely on pixel errors between generated and real images as the objective function resulted in relatively blurry generated images. To enhance super-resolution performance further, Ledig et al. introduced $SRGAN$ [42]. Generative Adversarial Networks (GAN) was the pioneer in utilizing GAN for image super-resolution processing, integrating two objective functions: pixel errors between generated and real images and perceptual loss between them. Perceptual loss captures variations at the deep feature level between generated and real images, guaranteeing that the generated images not only closely align with real images in terms of pixel values but also excel in terms of visual perception. This underscores the significant role of perceptual loss in augmenting the quality of generated images.

Building on the methodologies discussed earlier, this study enriches infrared images by employing a lightweight FCN guided by perceptual distance to reduce the perceptual gap between enhanced and original images for image enhancement. Unlike conventional supervised learning techniques, this method does not necessitate real data for training, offering a self-supervised approach.

3. Materials and Methods

The algorithm presented in this paper for infrared image perceptual enhancement treats the task as an image transformation challenge. It utilizes a lightweight FCN to transform the original image into an enhanced version, leveraging the $NLPD$ between the enhanced and original images as the objective function, facilitating network training via a self-supervised strategy. The algorithm's overall framework is depicted in Figure 1. Initially, the training dataset f is processed to extract inherent supervisory information as training labels S. Subsequently, f is passed through the lightweight FCN to derive a transformed image I that matches the size of f. Following this, the labels S and the transformed image

I undergo normalized Laplace transforms to obtain their multiscale transforms $f(I)$ and $f(S)$. Ultimately, the training process of *FCN* is supervised by minimizing the distance between $f(I)$ and $f(S)$.

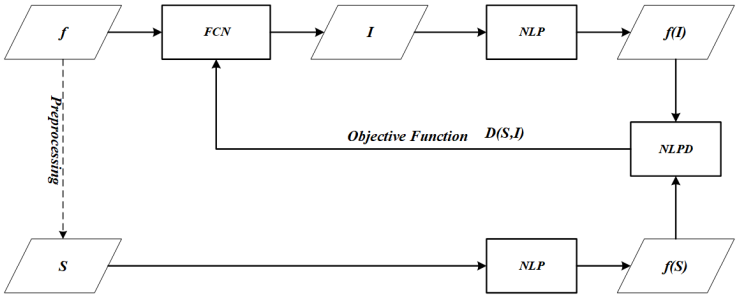

Figure 1. Overall framework of the proposed infrared enhancement method.

The actual effectiveness of the proposed infrared image perceptual enhancement algorithm is depicted in Figure 2. Compared to the original image and the method detailed in Section 2 of this paper, the image enhanced by the algorithm in this paper exhibits improved content and detail presentation, which is more conducive to human perception of the scene. The practical effect of the proposed infrared image perception enhancement algorithm is illustrated in Figure 2. Compared to the original image and methods detailed in Section 2 of this paper, *LIME* and *NPE* exhibit blurred and unclear branches, with missing architectural texture details and overall low image contrast. Images generated by *EnglightenGAN* show excessive noise, while *CLAHE* and *DHECL* exhibit artifacts and insufficient texture description, impacting visual perception. *ZeroDceP* and *Ours* proposed methods strike a good balance,effectively preserving crucial information and light details of targets like pedestrians, branches, and buildings. Among these methods, our proposed approach not only presents superior visual effects but also retains and enhances important target information and scene details.

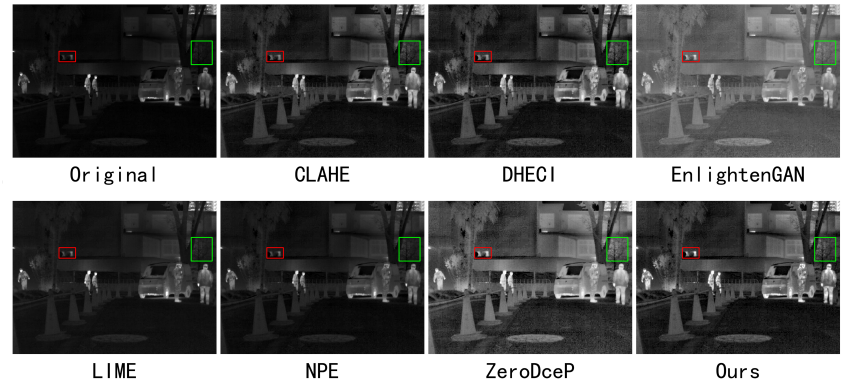

Figure 2. The actual effect of infrared image perception enhancement, and highlighting specific targets and points of interest using red and green boxes.

3.1. Lightweight FCN

We employ an *FCN* to achieve the transformation from input images to enhanced images, with the advantage of maintaining consistency in size between the input and output images. To ensure computational efficiency and speed of the algorithm, we adopt a lightweight *FCN*, aiming to minimize the number of network layers and parameter size while preserving the network's learning capabilities. The lightweight CNN network, with its simple yet effective three-layer architecture and the introduction of perceptual distance as a loss function, demonstrates excellent visual quality in infrared image enhancement. The images enhanced by this network not only exhibit more accurate details but also

present a more natural overall perception. Therefore, we propose the network architecture depicted in Figure 3.

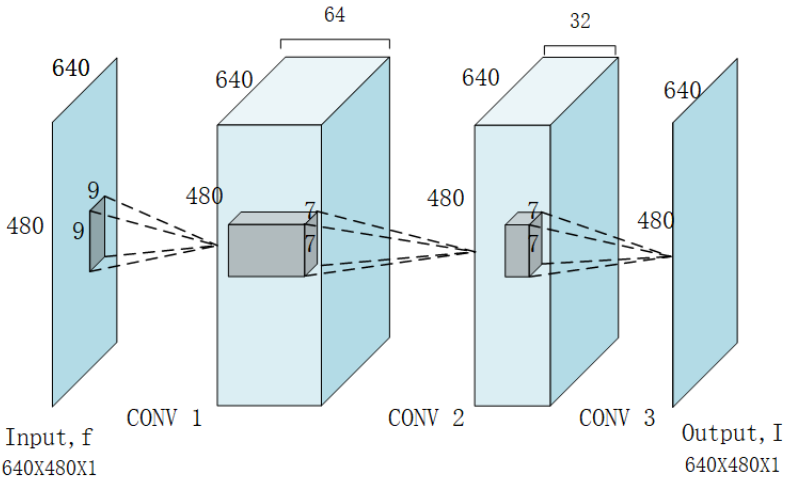

Figure 3. Lightweight *FCN* structure.

The network consists of three convolutional layers:

The first layer serves as the feature extraction layer, extracting overlapping image patches from the input image and mapping them to a high-dimensional space. This process is accomplished through convolutional filters. Formally, assuming the input image is denoted as f, the feature extraction operation can be expressed as follows:

$$F_1(f) = \sigma(\lambda_1 * f + B_1) \tag{1}$$

This layer takes the original image as input with 1 channel and produces image features with 64 channels. Each convolutional kernel has 9×9 weight parameters and 1 bias parameter. Therefore, this layer consists of a total of $64 \times (9 \times 9 + 1) = 5248$ parameters.

The second layer functions as the non-linear mapping layer, mapping high-dimensional feature representations to another set of high-dimensional feature representations. This non-linear mapping is crucial for capturing complex structures and patterns within image patches. Its expression is

$$F_2(F_1(f)) = \sigma(\lambda_2 * F_1(f) + B_2) \tag{2}$$

This layer takes image features as input with 64 channels and produces high-dimensional features with 32 channels. Each convolutional kernel has 9×9 weight parameters and 1 bias parameter. Therefore, this layer consists of a total of $32 \times ((64 \times 7 \times 7) + 1) = 100,384$ parameters.

The third layer serves as the image reconstruction layer, reconstructing the image from the high-dimensional features obtained through the non-linear mapping. Its expression is

$$I = \lambda_3 * F_2(F_1(f)) + B_3 \tag{3}$$

This layer takes high-dimensional features as input with 32 channels and produces the enhanced image with 1 channel. Each convolutional kernel has 7×7 weight parameters and 1 bias parameter. Therefore, this layer consists of a total of $32 \times 7 \times 7 + 1 = 1569$ parameters.

In summary, this network has a total of 107,201 parameters, making it a lightweight *FCN*.

In Equations (1)–(3), $*$ denotes the convolution operation, $\sigma()$ represents the activation function, $\lambda_1, \lambda_2, \lambda_3$ are convolutional kernels, and B_1, B_2, B_3 are bias terms. Here, I denotes the enhanced image output by this network.

Infrared images reflect the thermal radiation distribution of a scene. *CNN* can extract key information such as hotspots, edges, and textures from infrared images. Infrared images

are typically blurry and contain various types of noise, which hinder human recognition of targets and perception of the environment. By utilizing our proposed lightweight FCN, it is possible to effectively suppress noise in infrared images, enhance image details, improve image clarity, and enhance perceptual effects.

3.2. Objective Function Based on Metric Perceptual Distance

We employ the $NLPD$ to quantify the perceptual variances between images. $NLPD$ is intricately linked to the early human visual system; it initiates by constructing the Laplacian pyramid of an image [43], executes local luminance subtraction across various scales, and performs amplitude segmentation via local amplitude to diminish redundancy compared to the original image pixels. Substantial research suggests that this metric aligns more closely with human perception [44,45].

Assuming S represents the infrared radiation intensity of the real scene, $f(S)$ represents the perceptual result of humans towards S; I represents the displayed infrared image, and $f(I)$ represents the perceptual result of humans towards I. The intensity range of real-world infrared radiation is typically very large, while the brightness range of a display is usually between 5 and 300 (cd/m^2). This can lead to differences between the perceptual results $f(S)$ and $f(I)$ for humans. $NLPD$ can reflect the differences between the perceptual results $f(S)$ and $f(I)$.

3.2.1. NLP

Before subjecting the image S to NLP processing, preprocessing is essential. Initially, the image S undergoes a non-linear transformation grounded in visual biological principles, approximating the light response transformation in the photoreceptors of the retina [46], as follows:

$$X^{(0)} = S^r \tag{4}$$

Next, the preprocessed results undergo Gaussian pyramid transformation [47], as follows:

$$X^{(k+1)} = D(L(X^{(k)})),\ k = 0, 1, 2, \ldots, N \tag{5}$$

Then, the Laplacian pyramid $Z^{(k)}$ [48] is obtained, as follows:

$$Z^{(k)} = X^{(k)} - L(U(X^{(k+1)})),\ k = 0, 1, 2, \ldots, N \tag{6}$$

Finally, the normalized Laplacian pyramid $Y^{(k)}$ is obtained as follows:

$$Y^{(k)} = \frac{Z^{(k)}}{\sigma + p(|z^k|)},\ k = 0, 1, 2, \ldots, N \tag{7}$$

In the above expression, $D(\cdot)$ and $U(\cdot)$ represent linear downsampling and linear upsampling, respectively. L is a low-pass filter, k denotes the k-th level of the pyramid, the constant σ is a parameter for algorithm stability. P is a filter, commonly using the following template:

$$P = \begin{bmatrix} 0.04 & 0.04 & 0.05 & 0.04 & 0.04 \\ 0.04 & 0.03 & 0.04 & 0.03 & 0.04 \\ 0.05 & 0.04 & 0.05 & 0.04 & 0.05 \\ 0.04 & 0.03 & 0.04 & 0.03 & 0.04 \\ 0.04 & 0.04 & 0.05 & 0.04 & 0.04 \end{bmatrix} \tag{8}$$

3.2.2. Using $NLPD$ as the Loss Function

We use $NLPD$ to describe the difference between the original image S and the rendered image I, and we represent its image enhancement as a constrained optimization problem, aiming to optimize the rendered image I by minimizing the perceptual difference between the rendered image I and the original image S, as follows:

$$\hat{I} = \arg_I \min D(S, I) \tag{9}$$

where $D(,)$ is a measure of dissimilarity in human perception.

Perceptual difference is quantitatively represented by two parts, firstly, defining a non-linear perceptual transformation $f()$ that approximates the early processing of the human visual system. We apply this transformation to the original scene brightness S and the rendered image I, then measure the distance between $f(S)$ and $f(I)$.

The determination of the non-linear perceptual transformation $f()$:

From Equation (5), the combination of NLP coefficients $f(S)$ and $f(I)$ for each channel represents the response of the perceptual transformation, as follows:

$$f(S) = \{Y^{(k)} : k = 1, \ldots, N\} \tag{10}$$

$$f(I) = \{\tilde{Y}^{(k)} : k = 1, \ldots, N\} \tag{11}$$

Using this to measure the distance between $f(S)$ and $f(I)$, as follows:

$$D(S, I) = \left[\frac{1}{N} \sum_{k=1}^{N} \left(\frac{1}{N_c{}^{(k)}} \sum_{i=1}^{N_c} |Y_i^{(k)} - \tilde{Y}_i^{(k)}|^{\alpha} \right)^{\frac{\beta}{\alpha}} \right]^{\frac{1}{\beta}} \tag{12}$$

Therefore, we use Equation (12) as the $NLPD$ to describe the difference between the original image S and the rendered image I, serving as a similarity measure and supervising our lightweight CNN. Here, $Y_i^{(k)}$ and $\tilde{Y}^{(k)}$ represent corresponding layers of the normalized Laplacian pyramid, and α and β are parameters optimized to match human perceptual assessments of image quality in image quality databases.

3.3. Self-Supervision

By using $NLPD$ as a perceptual metric, we make a reasonable experimental estimation of the radiation intensity range of the real scene S for the displayed image I and linearly rescale the radiation intensity measurements. This resulting image is used as ground truth in the training model. This approach enables the lightweight CNN to train an infrared perceptual enhancement model through self-supervised learning. This process is specifically demonstrated as follows:

$$S = T\{[f]^{\mu}\} \tag{13}$$

Here, f represents the input image. Taking μ as 2.2 indicates a power transformation for *gamma* correction, utilized to enhance the display effect of images or adapt to the characteristics of the display device. T linearly maps the *gamma*-corrected range of f to [1, 3000], which is the radiation intensity range subjectively estimated through numerous experiments by assessors.

4. Experimental Results and Analysis

To validate the proposed infrared image enhancement method, comprehensive experiments were carried out on various public datasets. This section commences by outlining the experimental setup, datasets used, and evaluation metrics employed. Subsequently, the efficacy of the proposed method is confirmed through ablation and comparative experiments, followed by assessing the processing efficiency.

4.1. Experimental Setup

The experiments were conducted on a computer equipped with an *NVIDIA GeForce RTX* 4070 Laptop GPU (NVIDIA, Santa Clara, CA, USA). Training samples were image blocks of size 640×480. The model underwent training for 1000 epochs with a batch size of 8. The Adam optimizer [49] was utilized with a learning rate of 10^{-5}. In Equation (4), set $\gamma = 1/2.6$; similarly, in Equation (8), set $\sigma = 0.17$, in Equation (12), setting the values of α and β to 2 and 0.6, respectively; and fixing the number of levels N in the Laplacian pyramid

to 6, aims to best explain human perceptual ratings of distorted images in a common database [50]. Specifically, we select these parameters to maximize the correlation between the average scores given by human observers and the distances calculated by our metric.

4.2. Dataset

In this study, the experiments were conducted using four datasets: $HIT-UAV$ [51], $MSRS$ [52], TNO [53], and $VIFB$ [54]. The training was carried out using 1083 infrared images from the $MSRS$ dataset. For performance evaluation, 361 images from the $MSRS$ dataset, 25 images from the TNO dataset, 290 images from the $HIT-UAV$ dataset, and 21 images from the $VIFB$ dataset (totaling 697 infrared images) were utilized as the test dataset.

The $HIT-UAV$ dataset is a well-known dataset in the field of drone vision research. It offers images and video sequences captured by drones, specifically designed for tasks like object detection, tracking, recognition, and localization. This dataset serves as a valuable resource for researchers and practitioners working on drone vision-related applications.

The $MSRS$ dataset predominantly focuses on traffic scenes, encompassing diverse objects such as cars, pedestrians, and bicycles in both daytime and nighttime settings. Moreover, an image-enhancement algorithm grounded in the dark channel prior is applied to enhance the contrast and signal-to-noise ratio of infrared images within this dataset.

The TNO dataset consists of 63 images showcasing a range of military and surveillance scenes at various resolutions. These images depict a diverse array of objects and targets set against different backgrounds, including rural and urban environments. This dataset offers a valuable collection for research in military and surveillance imaging applications.

The $VIFB$ dataset comprises 21 infrared images sourced from the internet and various tracking datasets. These images span different resolutions and encompass a variety of environments and conditions, including indoor, outdoor, low-light, and overexposed scenarios. This dataset provides a diverse set of images useful for exploring various challenges and scenarios in infrared imaging applications.

We have selected some raw data from the aforementioned four datasets for presentation, as shown in Figure 4. Specifically, $a1$ to $a5$ represent data from the $HIT-UAV$ dataset; $b1$ to $b5$ correspond to the $MSRS$ dataset; $c1$ to $c5$ depict data from the TNO dataset; and $d1$ to $d5$ showcase the $VIFB$ dataset.

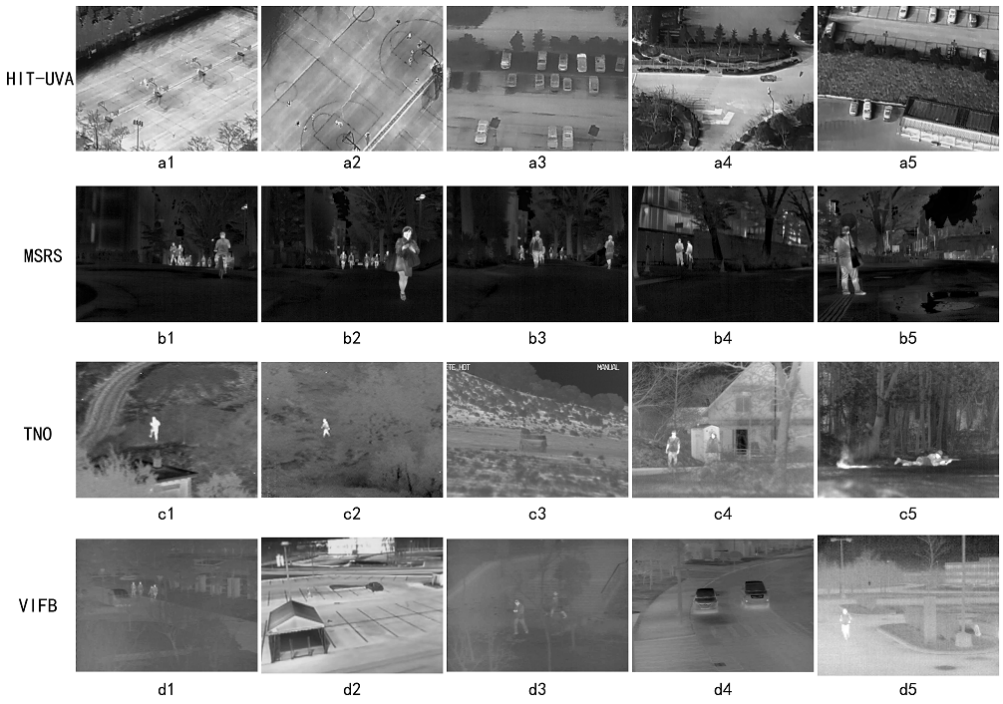

Figure 4. Display some sample images from the datasets used.

4.3. Evaluation Metrics

We evaluate the proposed infrared image enhancement method using six metrics: *NIQE* (Naturalness Image Quality Evaluator) [55], *EN* (Entropy) [56], *SSIM* , *AG* (Average Gradient) [57], *PSNR*, and *NLPD*.

NIQE is a metric crafted to evaluate the naturalness of an image by measuring how well the image quality aligns with human visual perception. A lower *NIQE* value indicates that the image closely resembles a natural image. The metric is defined as

$$\text{NIQE} = \sqrt{(\mu_1 - \mu_2)^T \Sigma^{-1} (\mu_1 - \mu_2)} \tag{14}$$

Among them, μ_1 and μ_2 represent the mean vectors of the evaluation image and the original natural image, while Σ denotes the covariance matrix.

EN is a metric utilized to measure the information content in an image, serving as a gauge of the image's complexity. A higher *EN* value signifies that the image contains a more extensive amount of information. This metric is defined as

$$\text{EN} = \sum_i P_i \log_2 P_i \tag{15}$$

Among them, P_i represents the probability of each gray level appearing in the image.

SSIM compares the structural similarity between two images by taking into account factors such as contrast, brightness, and local patterns. The definition of *SSIM* is as follows:

$$\text{SSIM}(x, y) = \frac{(2\mu_x\mu_y + C_1)}{(\mu_x^2 + \mu_y^2 + C_1)} \cdot \frac{(2\sigma_{xy} + C_2)}{(\sigma_x^2 + \sigma_y^2 + C_2)} \tag{16}$$

In the *SSIM* formula, x and y represent the source image and the enhanced image within the sliding window, respectively. σ_{xy} denotes the image covariance, while σ_x and σ_y represent the standard deviations of the images. μ_x and μ_y stand for the mean values of the images, respectively. C_1 and C_2 are parameters included to ensure algorithm stability. The *SSIM* value ranges from 0 to 1, where 1 indicates complete similarity between the two images.

AG is a metric employed to assess the sharpness of an image by calculating the average gradient of the image. A higher *AG* value suggests that the image is sharper. It is defined as

$$\text{AG} = \frac{1}{MN} \sum_{i=1}^{M} \sum_{j=1}^{N} \sqrt{\frac{(\nabla F_x(i,j))^2 + (\nabla F_y(i,j))^2}{2}} \tag{17}$$

In this context, ∇F_x and ∇F_y symbolize the gradient of the image in the x and y directions, respectively, while M and N denote the total number of pixels in the image.

The quality of an image can be evaluated by computing the *PSNR*, which represents the ratio of the signal to noise in the image. A higher *PSNR* value signifies better image quality. The formula to calculate *PSNR* is as follows:

$$\text{PSNR} = 10 \log_{10} \left(\frac{\text{MAX}^2}{\text{MSE}} \right) \tag{18}$$

In the *PSNR* formula provided earlier, *MAX* represents the maximum pixel value in the image, while *MSE* (Mean Squared Error) denotes the average squared difference between the original image and the processed image. The calculation formula for *MSE* is

$$\text{MSE} = \frac{1}{mn} \sum_{i=1}^{m} \sum_{j=1}^{n} (x(i,j) - y(i,j))^2 \tag{19}$$

In the context of the *PSNR* formula, $x(i,j)$ and $y(i,j)$ represent the pixel values of the original and processed images at coordinate (i,j), and m and n stand for the width and height of the image, respectively.

NLPD quantifies the disparity between the original scene and the enhanced image by evaluating the *NLPD* between them. A lower *NLPD* value signifies a higher quality of the processed image. It is defined as in Equation (12) above.

4.4. Ablation Study

In order to visually understand the effects of each major component of the perceptual loss metric, this section presents image renderings by selectively removing one of the three components of the perceptual loss metric. As shown in Figure 5, $a1$ to $a4$ display the original images; $b1$ to $b4$, $c1$ to $c4$, and $d1$ to $d4$, respectively, demonstrate the removal of one of the components of the perceptual loss metric: removal of the initial pointwise non-linearity (changing γ from 1/2.6 to 1 in Equation (4)), removal of the multi-scale decomposition (changing the number of layers N from 6 to 1 in the Laplacian pyramid), and removal of the split normalization (changing σ from 0.17 to 1 in Equation (7), and changing P to $P = 0$ in Equation (8)). On the other hand, $e1$ to $e5$ showcase images rendered using the complete perceptual loss metric.

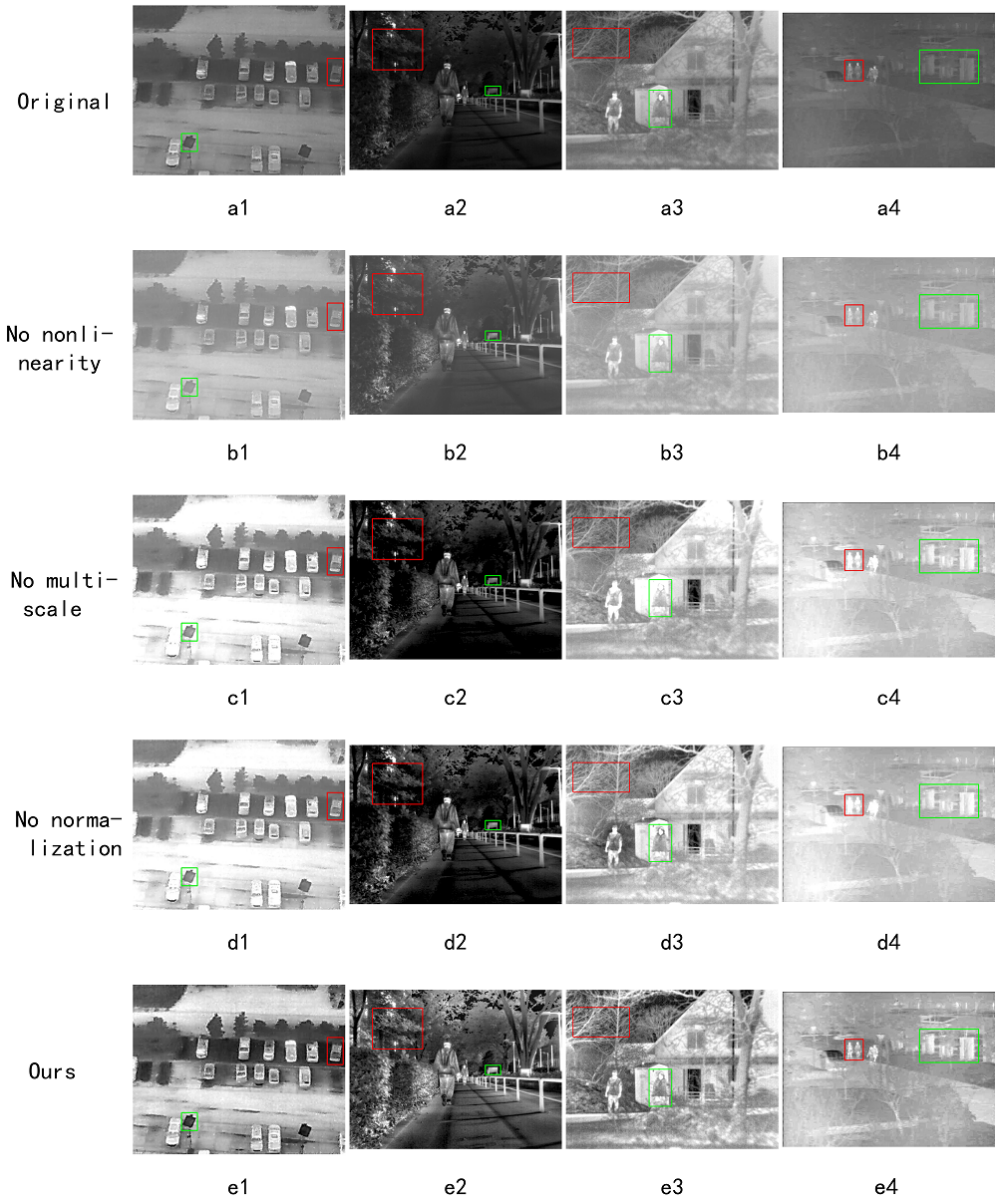

Figure 5. Visual comparison of ablation studies, and highlighting specific targets and points of interest using red and green boxes.

After removing the initial pointwise non-linearity component, the overall effect of the images becomes noticeably blurred with inadequate texture description, leading to compromised visual perception. Images where the multi-scale decomposition and split normalization components are removed exhibit severe color distortion or inconsistency, resulting in the loss of details in objects such as pedestrians and trees. Conversely, images rendered using the complete perceptual loss metric alleviate these issues and yield the most visually appealing results. For a closer examination of details, please refer to the ablation experiments shown in Figure 5.

4.5. Comparative Test Analysis

To validate the efficacy of our proposed method, we conducted a comparative analysis of its infrared enhancement performance against several methods detailed in Section 2 of this paper: $CLAHE$, $DHECL$, $EnglightenGAN$, HE, $LIME$, NPE, and $ZeroDceP$. To ensure fairness, default parameters provided by the respective authors were utilized for all comparison methods. These comparative experiments were carried out across four datasets, $HIT-UAV$, $MSRS$, TNO, and $VIFB$, encompassing qualitative evaluation, quantitative assessment, and subjective analysis. The qualitative evaluation involved visually inspecting the enhanced images, with specific targets and regions of interest highlighted using red and green boxes.

4.5.1. Results On The $HIT-UAV$ Dataset

In qualitative comparisons, $DHECL$, $EnglightenGAN$, and $LIME$ demonstrate significant limitations in preserving detailed textures and visual perception in the resultant images. In Figure 6a, $DHECL$ effectively enhances the aircraft visually but causes severe distortion in the ground near the aircraft. $LIME$ exhibits overall ground distortion and poor effects on multiple targets, leading to a significant loss of detail. In Figure 6a–c, $EnglightenGAN$ yields overall blurriness, accompanied by aliasing artifacts, inadequate texture description, and weak visual perception. Moreover, while NPE and $ZeroDceP$ can effectively display most targets in the enhanced images, they still lack details when representing certain small targets, such as the street lamp in Figure 6b. $ZeroDceP$ exhibits an overall high contrast in Figure 6, losing texture quality. Although many targets are clearly displayed, the enhancement effect on these targets is, at best, average. HE produces images with superior effects; however, it has some drawbacks. For instance, in the case of the aircraft and trees in Figure 6a, the street lamp in Figure 6b, and the wall in Figure 6c, the enhancement effects are slightly inferior to those generated by our proposed method. Our method effectively enhances and retains both the saliency information of targets and fine textures.

Quantitative Comparison: For a quantitative evaluation, 20 images from the $HIT-UAV$ dataset were utilized to compare the proposed method against six other enhancement techniques. The average results on the $HIT-UAV$ dataset are summarized in Table 1. Six metrics were employed to gauge the quality of the enhanced infrared images produced by the various methods. Our proposed method achieved the second-best results in the AG and EN evaluation metrics. Moreover, it notably secured the top position in the $NLPD$ metric, significantly surpassing the second-best method. A higher EN value indicates richer information content in the images, while a higher AG value signifies enhanced image clarity. Conversely, a lower $NLPD$ value suggests that the perceptual effect of the image aligns more closely with human perception.

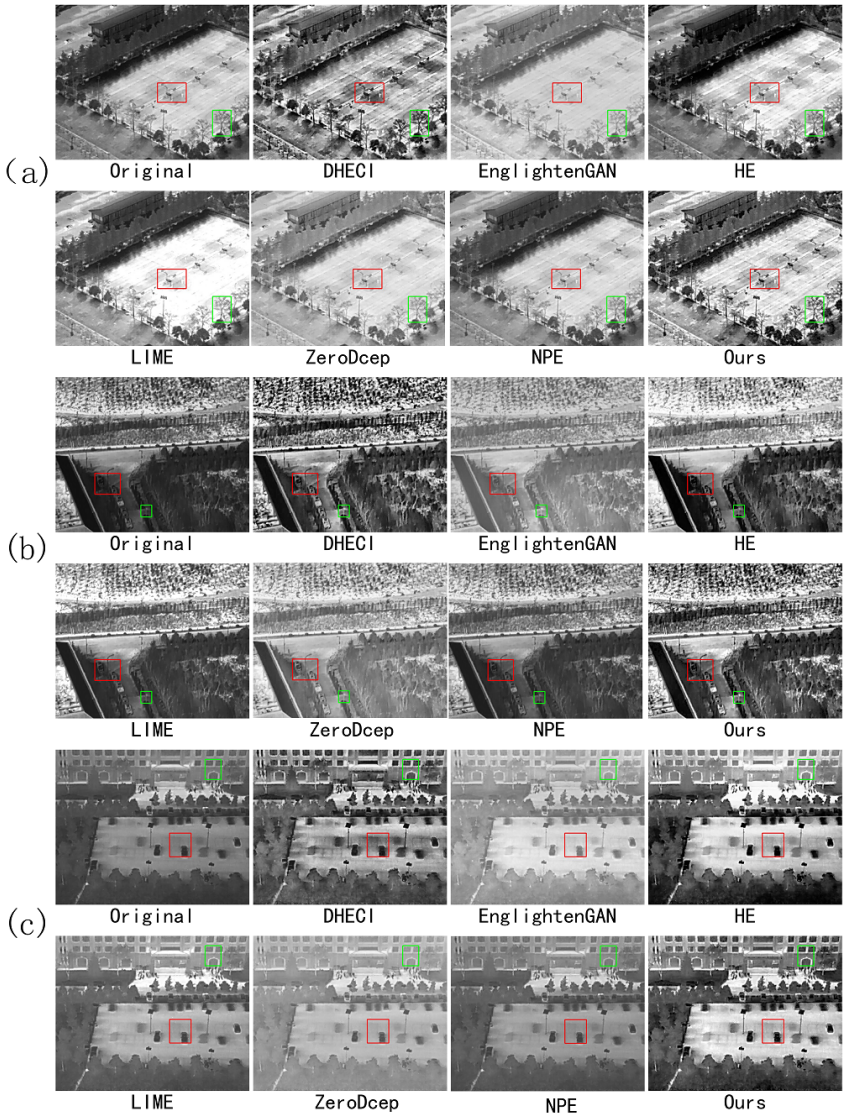

Figure 6. Qualitative comparison of three images (**a**–**c**) from the $HIT-UAV$ dataset with different algorithms. Provide detailed annotations using red and green boxes to emphasize important information.

Table 1. Quantitative results of comparative experiments on the $HIT-UAV$ dataset.

	NIQE	**EN**	**PSNR**	**AG**	**SSIM**	**NLPD**
DHECTI	5.0153	7.6206	17.8035	10.2708	0.7591	0.2220
EnglightenGAN	5.6981	6.9246	12.2964	4.6403	0.8102	0.2415
HE	5.0049	7.9568	19.0419	7.3519	0.8302	0.2086
LIME	5.5431	7.1541	18.1563	5.8662	50.9384	0.2136
NPE	5.9102	6.9183	23.3792	5.0112	0.9730	0.2321
ZeroDceP	5.5286	6.6832	12.7426	5.1908	0.8295	0.2411
Ours	5.2228	7.7903	17.2489	9.9996	0.79620	0.1236

4.5.2. Results on the *MSRS* Dataset

Qualitative Comparison: Figure 7 showcases the enhanced images produced by various algorithms in diverse scenarios. In these images, *LIME* and *NPE* present blurry tree branches with unclear structures, lacking texture details in buildings, inadequate texture description, and an overall low image contrast. *EnglightenGAN*'s images demonstrate excessive noise, impacting visual perception negatively. Conversely, *CLAHE*, *DHECL*, *ZeroDceP*, and our proposed method strike a good balance in various scenes, effectively preserving essential information about targets and lighting conditions like pedestrians, tree branches, and buildings. Among these methods, our proposed approach not only delivers superior visual effects but also preserves and enhances critical target information and scene details.

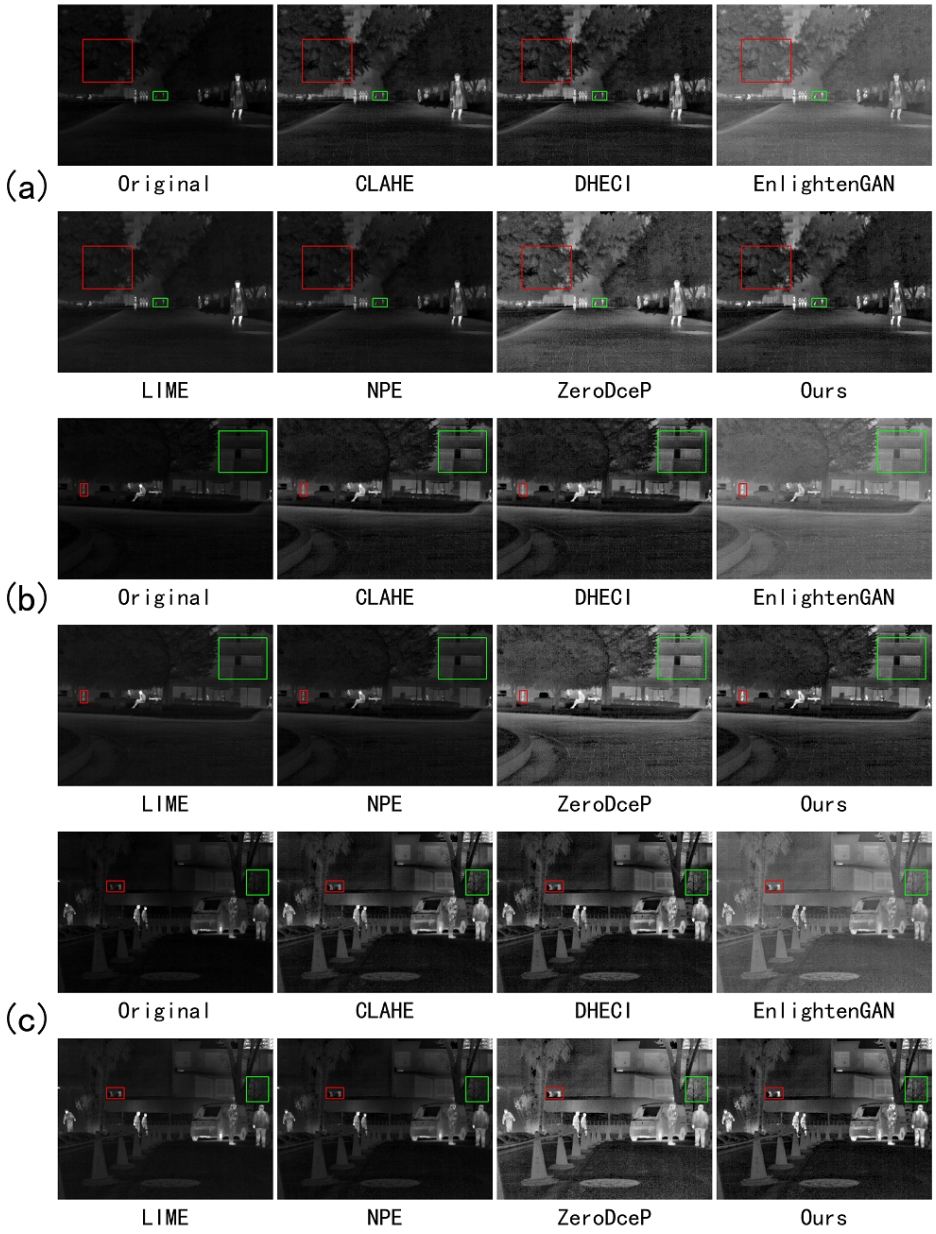

Figure 7. Qualitative comparison of three images (**a**–**c**) from the *MSRS* dataset, with some objects and details annotated with red and green boxes to highlight noteworthy information.

Quantitative Comparison: The proposed method has been quantitatively compared with six other enhancement methods on the *MSRS* dataset, and the average experimental

results on the *MSRS* dataset are presented in Table 2. Our proposed method excels in the *AG* and *EN* metrics, securing the second-best results, and attains the top position in the *NLPD* and *NIQE* metrics.

Table 2. Quantitative results of comparative experiments on the *MSRS* dataset.

	NIQE	EN	PSNR	AG	SSIM	NLPD
DHECTI	3.5195	6.7873	17.6502	4.8044	0.4899	0.0998
EnglightenGAN	3.4975	6.4964	9.0982	3.7535	0.2348	0.1020
CLAHE	3.6884	6.5106	25.8089	1.8769	0.7868	0.1339
LIME	4.8480	5.3585	23.8129	1.8196	0.7469	0.2686
NPE	4.4519	5.3609	27.5203	2.1749	0.8453	0.2291
ZeroDceP	3.4600	7.2106	11.8708	5.9975	0.2544	0.1224
Ours	3.4336	6.8434	15.8677	5.2749	0.4058	0.0535

4.5.3. Results on the *TNO* Dataset

Qualitative Comparison: The images generated by *EnglightenGAN* and *ZeroDceP* exhibit excessively high contrast, leading to detail loss and poor perceptual effects. Specifically, in Figure 8a, distinguishing between the fence and the ground outside is challenging. In Figure 8b, the brightness of the aircraft is excessively high, leading to a diminished contrast between the base of the trees and the branches, resulting in an overall lack of image clarity. Furthermore, in Figure 8c, the tree branches appear blurry. The images generated by *HE* exhibit noticeable aliasing artifacts and insufficient texture description. Images produced by *LIME* and *NPE* mildly enhance the original images, retaining complete details and textures, but the enhancement effect is relatively weak. Conversely, *DHECL*, while preserving details and textures intact, offers superior visual perception enhancements in different aspects. Nonetheless, in terms of visual effects, our proposed method still attains the best results.

Quantitative Comparison: The proposed method has been quantitatively compared with six other enhancement methods using images from the *TNO* dataset. The average experimental results on the *TNO* dataset are detailed in Table 3. Our proposed method excels in the *NIQE* and *AG* metrics and secures the top position in the *NLPD* and *EN* metrics.

Table 3. Quantitative results of comparative experiments on the *TNO* dataset.

	NIQE	EN	PSNR	AG	SSIM	NLPD
DHECTI	6.4298	6.8943	19.6723	8.2800	0.7058	0.1652
EnglightenGAN	6.2767	5.9904	10.3586	3.6241	0.8146	0.2086
HE	6.0037	6.0579	13.5681	8.9784	0.5819	0.2401
LIME	6.8526	5.7310	18.4287	3.4361	0.9558	0.2166
NPE	7.2304	6.0679	22.7163	2.8657	0.9821	0.2468
ZeroDceP	7.0859	5.9203	12.7361	3.0101	0.8799	0.2448
Ours	6.4191	7.4784	13.4904	7.4619	0.7242	0.0858

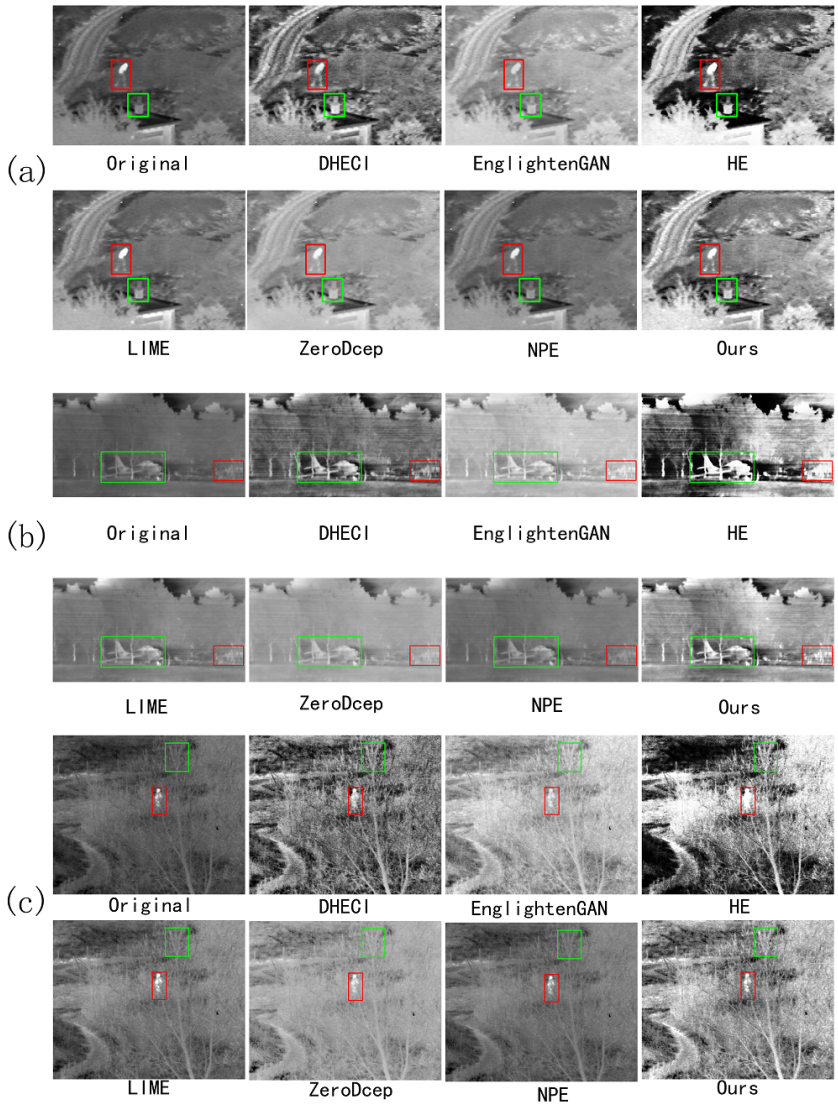

Figure 8. Qualitative comparison of three images (**a**–**c**) from the *TNO* dataset, with some objects and details annotated with red and green boxes to highlight noteworthy information.

4.5.4. Results on the *VIFB* Dataset

Qualitative Comparison: The images produced by *HE* exhibit noticeable artifacts and lack sufficient texture description. For instance, in Figure 9a, the ground, in Figure 9b, the building, and in Figure 9c, the road suffer from these issues. Furthermore, *HE* overexposes the wall in Figure 9a, resulting in an overall low image quality. The images produced by *EnglightenGAN*, *LIME*, *NPE*, and *ZeroDceP* in Figure 9a are of poor quality, appearing blurry with unclear target boundaries and insufficient detail description. The contrast in the images generated by *EnglightenGAN* and *ZeroDceP* in Figure 9a,b is excessively high, resulting in poor visual quality. The image enhancement effects of *LIME*, *NPE*, and *ZeroDceP* in Figure 9b,c are average, as they struggle to enhance the details of their targets effectively. *DHECI* and our proposed method effectively preserve and enhance the crucial information present in the original images, including pedestrians, branches, buildings, vehicles, zebra crossings, and more. Notably, our proposed method excels in capturing details and aligns closely with human perception, distinguishing it as a standout performer in the image enhancement process.

Quantitative Comparison: The proposed method has undergone a quantitative comparison with six other enhancement methods using the *VIFB* dataset. The average experimental results on the *VIFB* dataset are provided in Table 4. Our proposed method

demonstrates strong performance across the *NIQE*, *EN*, and *AG* metrics and secures the top position in terms of absolute performance in the *NLPD* indicator.

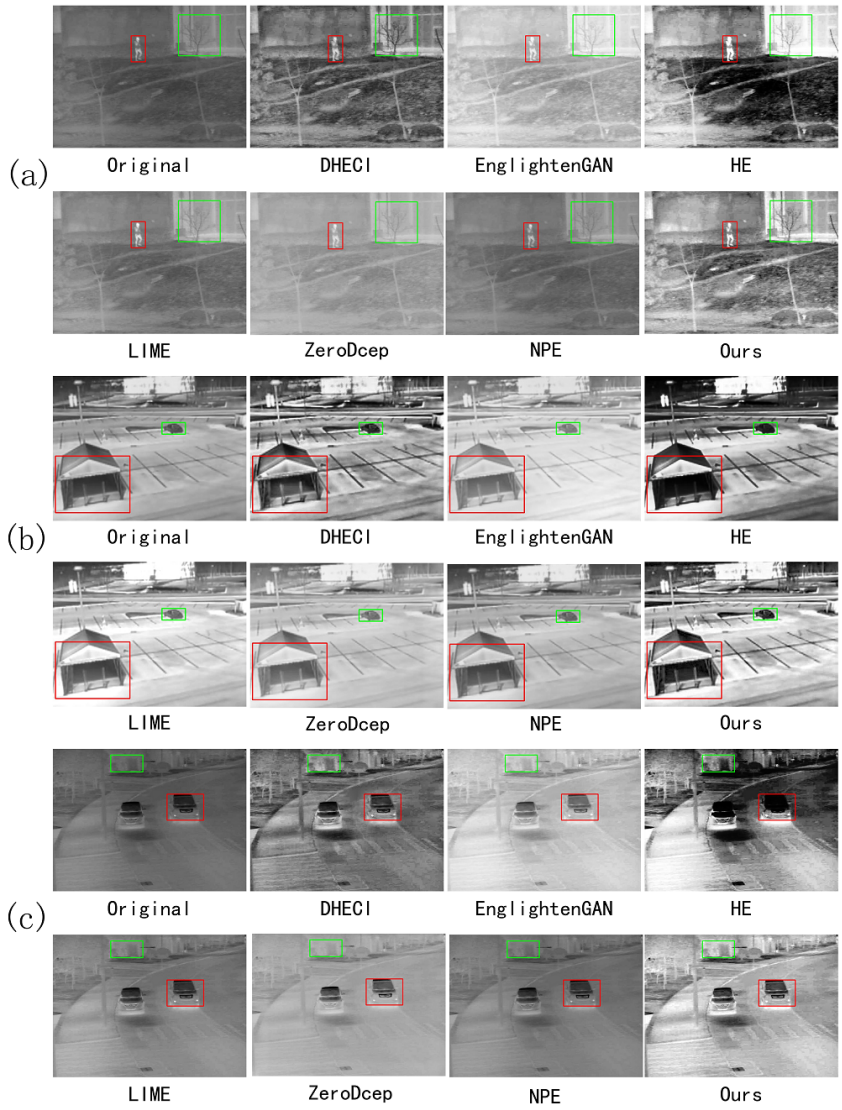

Figure 9. Qualitative comparison of three images (**a**–**c**) from the *VIFB* dataset, with some objects and details annotated with red and green boxes to highlight noteworthy information.

Table 4. Quantitative results of comparative experiments on the *VIFB* dataset.

	NIQE	EN	PSNR	AG	SSIM	NLPD
DHECTI	4.7578	6.8732	20.8278	5.7524	0.7574	0.1656
EnglightenGAN	5.6615	6.3631	10.5073	2.6205	0.8274	0.1905
HE	4.4711	7.8325	13.5032	6.0706	0.6356	0.2348
LIME	5.7914	6.4177	17.7937	2.5862	0.9585	0.2272
NPE	6.0682	6.1399	23.0080	2.0974	0.9835	0.2390
ZeroDceP	5.9848	5.6649	13.5126	2.1087	0.8981	0.2341
Ours	4.8693	7.0655	13.7773	5.9834	0.7154	0.0891

4.5.5. People's Subjective Evaluation

In our study, we conducted a human subjective evaluation to compare the performance of our proposed infrared perceptual enhancement method with other methods. We randomly selected five images from each of the test sets $HIT-UAV$, $MSRS$, TNO, and $VIFB$, totaling 20 images. For each image, we applied seven different methods ($DHECI$, $EnglightenGAN$, HE, $LIME$, NPE, $ZeroDceP$, $Ours$) to enhance it. Subsequently, we presented these seven output images to eleven participants for evaluation. During the evaluation, participants were shown a randomly selected pair of images from the seven outputs and asked to determine which image exhibited better quality in each comparison.

The quality was evaluated based on the following criteria:

(1) Whether the image exhibited texture distortion;
(2) Whether the image contained visible noise;
(3) Whether the image contained over-exposed or under-exposed artifacts.

By having participants subjectively score the 7 methods, we were able to rank the methods from 1 to 7 based on their perceived quality. This ranking process was repeated for all 20 images, and the results are displayed in Figure 10.

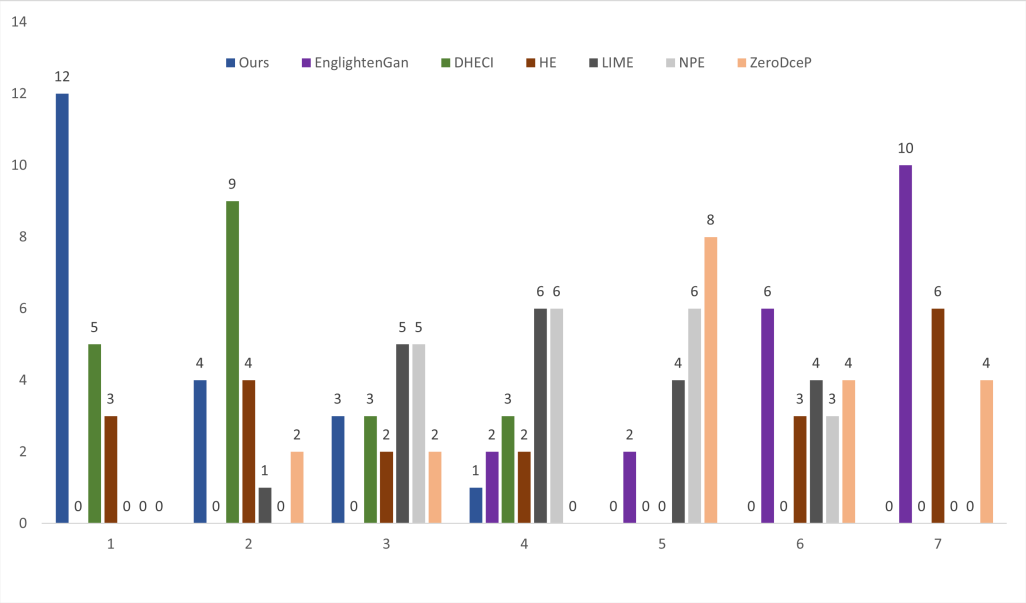

Figure 10. The results of the human subjective evaluation for the 7 enhancement methods are presented in the form of histograms. In each histogram, the x-axis represents the ranking levels (1∼7, with 1 being the highest rank), and the y-axis indicates the number of images assigned to each ranking level.

In Figure 10, the seven histograms depict the distribution of overall scores given by participants for the seven enhancement methods across the 20 test images. A comparison of these histograms clearly indicates that our proposed method yielded the most favorable results, according to the human participants. $EnglightenGAN$ received the lowest scores, likely due to its tendency to cause overexposure and sometimes amplify noise. HE and $ZeroDceP$ received varying scores, with HE and $ZeroDceP$ being suitable for some scene images but less effective for others. $LIME$ and NPE received average scores, indicating moderate overall image enhancement effects and inadequate detailed descriptions.

4.5.6. Computational Efficiency Analysis

To evaluate processing efficiency, we conducted 10 runs of all images within the $HIT-UAV$, $MSRS$, TNO, and $VIFB$ datasets using our proposed method and the other six comparison methods, following the previously mentioned configuration. Subsequently,

we calculated the average processing time for each method, and the outcomes are summarized in Table 5. The data showcase that our proposed method achieved the fastest average processing speed in comparison to existing image enhancement techniques. This underscores the computational efficiency of our approach, a critical aspect for practical applications that necessitate real-time or near-real-time performance.

Table 5. Each method processes the results of the average computational efficiency of the image.

Time (seconds)	HIT-UAV	MSRS	TNO	VIFB
DHECTI	0.0225	0.0316	0.0292	0.0390
EnglightenGAN	0.0276	0.0337	0.0266	0.0323
HE	0.0091	0.0195	0.0178	0.0228
LIME	0.0197	0.0298	0.0235	0.0259
NPE	0.0115	0.0185	0.0229	0.0217
ZeroDceP	0.0308	0.0327	0.0366	0.0365
Ours	0.0073	0.0122	0.0098	0.0177

5. Discussion

Based on the experimental results and analysis, our proposed method has showcased superior subjective performance when compared to the seven comparison methods across the four test datasets, underscoring the effectiveness and robustness of our approach. While our method may not surpass others in certain objective metrics such as $SNRP$ and $SSIM$, it excels in providing a more balanced enhancement of image quality across various aspects. In terms of other objective measures, our method demonstrates relatively better performance compared to the comparison techniques. Particularly noteworthy is its significant outperformance in the $NLPD$ metric, which closely aligns with human perceptual assessment. Moreover, our method boasts the fastest processing speed among current infrared image enhancement techniques, highlighting its computational efficiency as a key advantage.

6. Conclusions

In this research paper, we introduced a novel, straightforward, and efficient approach based on an FCN for self-supervised perceptual enhancement of infrared images, aiming to enhance images in a manner that aligns closely with human perception. Our method incorporates the $NLPD$ metric to evaluate enhancement effects, bridging the gap between objective metric assessments and visual perceptual mechanisms, effectively. Additionally, our approach has showcased the fastest computational efficiency in processing enhanced infrared images, compared to the methods examined in our study. For future work, we intend to delve into the realm of infrared and visible light image fusion using a perceptual loss metric. We also plan to broaden the scope of our method's evaluation to encompass a wider array of application scenarios. Furthermore, we aim to explore how perceptual enhancement techniques can enhance the efficacy of various visual tasks such as object detection, tracking, and segmentation.

Author Contributions: Methodology, Z.Z. and Y.X.; writing advice, Z.Z. and Z.L. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Data Availability Statement: The original data presented in the study are openly available at https://github.com/suojiashun/HIT-UAV-Infrared-Thermal-Dataset; accessed on 7 September 2024.

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

References

1. Fan, Z.; Bi, D.; Xiong, L.; Ma, S.; He, L.; Ding, W. Dim infrared image enhancement based on convolutional neural network. *Neurocomputing* **2018**, *272*, 396–404. [CrossRef]
2. Kuang, X.; Sui, X.; Liu, Y.; Chen, Q.; Gu, G. Single infrared image enhancement using a deep convolutional neural network. *Neurocomputing* **2019**, *332*, 119–128. [CrossRef]
3. He, K.; Zhang, X.; Ren, S.; Sun, J. Deep residual learning for image recognition. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 27–30 June 2016; pp. 770–778.
4. Siddique, N.; Paheding, S.; Elkin, C.P.; Devabhaktuni, V. U-net and its variants for medical image segmentation: A review of theory and applications. *IEEE Access* **2021**, *9*, 82031–82057. [CrossRef]
5. Gao, Z.; Dai, J.; Xie, C. Dim and small target detection based on feature mapping neural networks. *J. Vis. Commun. Image Represent.* **2019**, *62*, 206–216. [CrossRef]
6. Han, J.; Moradi, S.; Faramarzi, I.; Zhang, H.; Zhao, Q.; Zhang, X.; Li, N. Infrared small target detection based on the weighted strengthened local contrast measure. *IEEE Geosci. Remote. Sens. Lett.* **2020**, *18*, 1670–1674. [CrossRef]
7. Yuan, Z.; Jia, L.; Wang, P.; Zhang, Z.; Li, Y.; Xia, M. Infrared Image Enhancement Based on Multiple Scale Retinex and Sequential Guided Image Filter. In Proceedings of the 2024 3rd Asia Conference on Algorithms, Computing and Machine Learning, Shanghai, China, 22–24 March 2024; pp. 196–201.
8. Shanmugavadivu, P.; Balasubramanian, K. Particle swarm optimized multi-objective histogram equalization for image enhancement. *Opt. Laser Technol.* **2014**, *57*, 243–251. [CrossRef]
9. Gupta, B.; Tiwari, M. A tool supported approach for brightness preserving contrast enhancement and mass segmentation of mammogram images using histogram modified grey relational analysis. *Multidimens. Syst. Signal Process.* **2017**, *28*, 1549–1567. [CrossRef]
10. Huang, S.C.; Cheng, F.C.; Chiu, Y.S. Efficient contrast enhancement using adaptive gamma correction with weighting distribution. *IEEE Trans. Image Process.* **2012**, *22*, 1032–1041. [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. Wei, C.; Wang, W.; Yang, W.; Liu, J. Deep retinex decomposition for low-light enhancement. *arXiv* **2018**, arXiv:1808.04560.
13. Petro, A.B.; Sbert, C.; Morel, J.M. Multiscale retinex. *Image Process. Line* **2014**, *4*, 71–88. [CrossRef]
14. Yang, W.; Wang, W.; Huang, H.; Wang, S.; Liu, J. Sparse gradient regularized deep retinex network for robust low-light image enhancement. *IEEE Trans. Image Process.* **2021**, *30*, 2072–2086. [CrossRef]
15. 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]
16. Levoy, M.; Hanrahan, P. Light field rendering. In *Seminal Graphics Papers: Pushing the Boundaries*; ACM Digital Library: New York, NY, USA, 2023; Volume 2, pp. 441–452.
17. Gong, X.; Chang, S.; Jiang, Y.; Wang, Z. Autogan: Neural architecture search for generative adversarial networks. In Proceedings of the IEEE/CVF International Conference on Computer Vision, Seoul, Republic of Korea, 27 October–2 November 2019; pp. 3224–3234.
18. Gong, C.; Tao, D.; Maybank, S.J.; Liu, W.; Kang, G.; Yang, J. Multi-modal curriculum learning for semi-supervised image classification. *IEEE Trans. Image Process.* **2016**, *25*, 3249–3260. [CrossRef]
19. Rani, V.; Nabi, S.T.; Kumar, M.; Mittal, A.; Kumar, K. Self-supervised learning: A succinct review. *Arch. Comput. Methods Eng.* **2023**, *30*, 2761–2775. [CrossRef] [PubMed]
20. Papandreou, G.; Chen, L.C.; Murphy, K.P.; Yuille, A.L. Weakly-and semi-supervised learning of a deep convolutional network for semantic image segmentation. In Proceedings of the IEEE International Conference on Computer Vision, Santiago, Chile, 7–13 December 2015; pp. 1742–1750.
21. Huang, X.; Liu, M.Y.; Belongie, S.; Kautz, J. Multimodal unsupervised image-to-image translation. In Proceedings of the European Conference on Computer Vision (ECCV), Munich, Germany, 8–14 September 2018; pp. 172–189.
22. Liu, M.Y.; Breuel, T.; Kautz, J. Unsupervised image-to-image translation networks. In Proceedings of the NIPS'17: Proceedings of the 31st International Conference on Neural Information Processing Systems, Long Beach, CA, USA, 4–9 December 2017.
23. Laparra, V.; Ballé, J.; Berardino, A.; Simoncelli, E.P. Perceptual image quality assessment using a normalized Laplacian pyramid. *Electron. Imaging* **2016**, *28*, 1–6. [CrossRef]
24. Laparra, V.; Berardino, A.; Ballé, J.; Simoncelli, E.P. Perceptually optimized image rendering. *JOSA A* **2017**, *34*, 1511–1525. [CrossRef] [PubMed]
25. Abdullah-Al-Wadud, M.; Kabir, M.H.; Dewan, M.A.A.; Chae, O. A dynamic histogram equalization for image contrast enhancement. *IEEE Trans. Consum. Electron.* **2007**, *53*, 593–600. [CrossRef]
26. Nakai, K.; Hoshi, Y.; Taguchi, A. Color image contrast enhacement method based on differential intensity/saturation gray-levels histograms. In Proceedings of the 2013 International Symposium on Intelligent Signal Processing and Communication Systems, Okinawa, Japan, 12–15 November 2013; pp. 445–449.
27. Reza, A.M. Realization of the contrast limited adaptive histogram equalization (CLAHE) for real-time image enhancement. *J. VLSI Signal Process. Syst. Signal Image Video Technol.* **2004**, *38*, 35–44. [CrossRef]

28. Wang, Y.; Cao, Y.; Zha, Z.J.; Zhang, J.; Xiong, Z.; Zhang, W.; Wu, F. Progressive retinex: Mutually reinforced illumination-noise perception network for low-light image enhancement. In Proceedings of the 27th ACM International Conference on Multimedia, Nice, France, 21–25 October 2019; pp. 2015–2023.
29. Guan, X.; Jian, S.; Hongda, P.; Zhiguo, Z.; Haibin, G. An image enhancement method based on gamma correction. In Proceedings of the 2009 Second International Symposium on Computational Intelligence and Design, Changsha, China, 12–14 December 2009; Volume 1, pp. 60–63.
30. Zhang, D.; Zhang, D. Wavelet transform. In *Fundamentals of Image Data Mining: Analysis, Features, Classification and Retrieval*; Springer: Berlin/Heidelberg, Germany, 2019; pp. 35–44.
31. Guo, X.; Li, Y.; Ling, H. LIME: Low-light image enhancement via illumination map estimation. *IEEE Trans. Image Process.* **2016**, *26*, 982–993. [CrossRef]
32. Fan, G.D.; Fan, B.; Gan, M.; Chen, G.Y.; Chen, C.P. Multiscale low-light image enhancement network with illumination constraint. *IEEE Trans. Circuits Syst. Video Technol.* **2022**, *32*, 7403–7417. [CrossRef]
33. Cha, D.; Jeong, S.; Yoo, M.; Oh, J.; Han, D. Multi-input deep learning based FMCW radar signal classification. *Electronics* **2021**, *10*, 1144. [CrossRef]
34. Wu, W.; Weng, J.; Zhang, P.; Wang, X.; Yang, W.; Jiang, J. Uretinex-net: Retinex-based deep unfolding network for low-light image enhancement. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, New Orleans, LA, USA, 18–24 June 2022; pp. 5901–5910.
35. Jiang, Y.; Gong, X.; Liu, D.; Cheng, Y.; Fang, C.; Shen, X.; Yang, J.; Zhou, P.; Wang, Z. Enlightengan: Deep light enhancement without paired supervision. *IEEE Trans. Image Process.* **2021**, *30*, 2340–2349. [CrossRef] [PubMed]
36. Liu, R.; Ma, L.; Zhang, J.; Fan, X.; Luo, Z. Retinex-inspired unrolling with cooperative prior architecture search for low-light image enhancement. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Nashville, TN, USA, 20–25 June 2021; pp. 10561–10570.
37. Li, C.; Guo, C.; Loy, C.C. Learning to enhance low-light image via zero-reference deep curve estimation. *IEEE Trans. Pattern Anal. Mach. Intell.* **2021**, *44*, 4225–4238. [CrossRef] [PubMed]
38. Setiadi, D.R.I.M. PSNR vs. SSIM: Imperceptibility quality assessment for image steganography. *Multimed. Tools Appl.* **2021**, *80*, 8423–8444. [CrossRef]
39. Hore, A.; Ziou, D. Image quality metrics: PSNR vs. SSIM. In Proceedings of the 2010 20th International Conference on Pattern Recognition, Istanbul, Turkey, 23–26 August 2010; pp. 2366–2369.
40. Walach, H.; Buchheld, N.; Buttenmüller, V.; Kleinknecht, N.; Schmidt, S. Measuring mindfulness—the Freiburg mindfulness inventory (FMI). *Personal. Individ. Differ.* **2006**, *40*, 1543–1555. [CrossRef]
41. Dong, C.; Loy, C.C.; Tang, X. Accelerating the super-resolution convolutional neural network. In Proceedings of the Computer Vision–ECCV 2016: 14th European Conference (Proceedings, Part II 14), Amsterdam, The Netherlands, 11–14 October 2016; pp. 391–407.
42. Ledig, C.; Theis, L.; Huszár, F.; Caballero, J.; Cunningham, A.; Acosta, A.; Aitken, A.; Tejani, A.; Totz, J.; Wang, Z.; et al. Photo-realistic single image super-resolution using a generative adversarial network. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, HI, USA, 21–26 July 2017; pp. 4681–4690.
43. Burt, P.J.; Adelson, E.H. The Laplacian pyramid as a compact image code. In *Readings in Computer Vision*; Elsevier: Amsterdam, The Netherlands, 1987; pp. 671–679.
44. Paris, S.; Hasinoff, S.W.; Kautz, J. Local Laplacian filters: Edge-aware image processing with a Laplacian pyramid. *ACM Trans. Graph.* **2011**, *30*, 68. [CrossRef]
45. Laparra, V.; Muñoz-Marí, J.; Malo, J. Divisive normalization image quality metric revisited. *JOSA A* **2010**, *27*, 852–864. [CrossRef]
46. Heeger, D.J. Normalization of cell responses in cat striate cortex. *Vis. Neurosci.* **1992**, *9*, 181–197. [CrossRef]
47. Lan, Z.; Lin, M.; Li, X.; Hauptmann, A.G.; Raj, B. Beyond gaussian pyramid: Multi-skip feature stacking for action recognition. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Boston, MA, USA, 7–12 June 2015; pp. 204–212.
48. Lai, W.S.; Huang, J.B.; Ahuja, N.; Yang, M.H. Deep laplacian pyramid networks for fast and accurate super-resolution. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, HI, USA, 21–26 July 2017; pp. 624–632.
49. Kingma, D.P.; Ba, J. Adam: A method for stochastic optimization. *arXiv* **2014**, arXiv:1412.6980.
50. Ponomarenko, N.; Lukin, V.; Zelensky, A.; Egiazarian, K.; Carli, M.; Battisti, F. TID2008-a database for evaluation of full-reference visual quality assessment metrics. *Adv. Mod. Radioelectron.* **2009**, *10*, 30–45.
51. Suo, J.; Wang, T.; Zhang, X.; Chen, H.; Zhou, W.; Shi, W. HIT-UAV: A high-altitude infrared thermal dataset for Unmanned Aerial Vehicle-based object detection. *Sci. Data* **2023**, *10*, 227. [CrossRef] [PubMed]
52. Serp, J.; Allibert, M.; Beneš, O.; Delpech, S.; Feynberg, O.; Ghetta, V.; Heuer, D.; Holcomb, D.; Ignatiev, V.; Kloosterman, J.L.; et al. The molten salt reactor (MSR) in generation IV: Overview and perspectives. *Prog. Nucl. Energy* **2014**, *77*, 308–319. [CrossRef]
53. Kuenen, J.; Visschedijk, A.; Jozwicka, M.; Denier Van Der Gon, H. TNO-MACC_II emission inventory; a multi-year (2003–2009) consistent high-resolution European emission inventory for air quality modelling. *Atmos. Chem. Phys.* **2014**, *14*, 10963–10976. [CrossRef]
54. Zhang, X.; Ye, P.; Xiao, G. VIFB: A visible and infrared image fusion benchmark. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops, Seattle, WA, USA, 14–19 June 2020; pp. 104–105.

55. Zhang, L.; Zhang, L.; Bovik, A.C. A feature-enriched completely blind image quality evaluator. *IEEE Trans. Image Process.* **2015**, *24*, 2579–2591. [CrossRef]
56. Guttman, A. R-trees: A dynamic index structure for spatial searching. In Proceedings of the 1984 ACM SIGMOD International Conference on Management of Data, Boston MA, USA, 18–21 June 1984; pp. 47–57.
57. Horn-von Hoegen, M.; Schmidt, T.; Meyer, G.; Winau, D.; Rieder, K. Lattice accommodation of low-index planes: Ag (111) on Si (001). *Phys. Rev. B* **1995**, *52*, 10764. [CrossRef]