Journal of Cloud Computing:
Advances, Systems and Applications

RESEARCH

Open Access

LAE-GAN: a novel cloud-based Low-light Attention Enhancement Generative Adversarial Network for unpaired text images

Minglong Xue[1,2], Yanyi He[1], Peiqi Xie[1], Zhengyang He[1] and Xin Feng[1*]

Abstract

With the widespread adoption of mobile multimedia devices, the deployment of compute-intensive inference tasks on edge and resource-constrained devices, particularly in the context of low-light text detection, remains a formidable challenge. Existing deep learning approaches have shown limited effectiveness in restoring images for extremely dark scenes. To address these limitations, this paper presents a novel cloud-based **L**ow-light **A**ttention **E**nhancement **G**enerative **A**dversarial **N**etwork for unpaired text images **(LAE-GAN)** for the non-paired text image enhancement task in extremely low-light conditions. In the first stage, compressed low-light images are transmitted from edge devices to a cloud server for image enhancement. The LAE-GAN, an end-to-end network comprising a Zero-DCE and AGM-net generator, is designed with a global and local discriminator structure. The initial illumination restoration of extremely low-light images is accomplished using the Zero-DCE network. To enhance text details, we propose an Enhanced Text Attention Mechanism (ETAM) that transforms text information into a comprehensive text attention mechanism across the entire network. The Sobel operator is employed to extract text edge information, while attention is focused on text region details through constraints imposed on the attention map and edge map. Additionally, an AGM-Net module is integrated to reduce noise and fine-tune illumination. In the second stage, the cloud server makes decisions based on user requirements and processes requests in parallel, scaling with the quantity of requests. In the third stage, the enhanced results are transmitted back to edge devices for text detection. Experimental results on widely used LOL and SID low-light datasets demonstrate significant improvements in both quantitative and qualitative analysis, surpassing state-of-the-art enhancement methods in terms of image restoration and text detection.

Keywords Low-light enhancement, Cloud computing, Unpaired image, LAE-GAN, Textual attention mechanism

Introduction

In the era of the Internet of Things [1–3], a solid technological foundation for smart cities has been laid with the popularity and development of connected devices [4–6], which make it easy to take photos and upload them to the cloud at any time [7–10]. However, the quality of the captured photos may be compromised due to environmental limitations and device constraints, particularly in low-light conditions. Additionally, with the rise of smart cities, there is an increasing demand for cameras that can accurately capture nighttime scenes. Therefore, obtaining high-quality images is crucial for our perception and understanding of the surrounding environment. Images captured in low-light scenes often exhibit characteristics such as low contrast, low saturation, and significant noise, making it highly challenging to directly understand the content of images captured under low-light conditions. Currently, most visual tasks are designed for

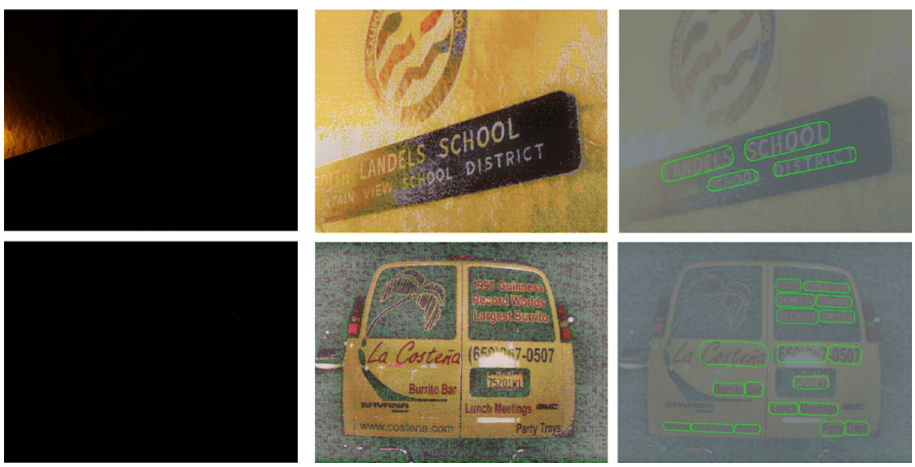
*Correspondence:
Xin Feng
xfeng@cqut.edu.cn
[1] School of Computer Science and Engineering, Chongqing University of Technology, Chongqing, China
[2] National Key Lab for Novel Software Technology, Nanjing University, Nanjing, China

non-low-light images, and their detection performance is poor in weak lighting conditions. However, due to inevitable environmental or technical constraints, many photos are often captured in low-light scenes. In addition, inadequate and unbalanced lighting conditions, as well as insufficient exposure during image capture, may be present. These low-light images suffer from compromised image quality and unfavorable information transmission, which not only affect viewers' visual experience but also result in conveying incorrect information, such as erroneous target recognition.

In recent years, with the advancement of deep learning, it has not only been extensively applied in other fields like natural language processing [11] and recommendation systems [12] but has also achieved significant accomplishments in the computer vision domain. Particularly, in the realm of low-light enhancement methods, deep learning has made remarkable progress, offering powerful solutions to improve image quality and performance in low-light conditions. These notable advancements have brought about breakthroughs and innovations in the computer vision field, driving continuous development in image processing and applications. Currently, deep learning-based enhancement methods heavily rely on paired data to train models for image restoration tasks such as super-resolution [13], denoising [14], and deblurring [15]. Overall, these methods have improved the quality of low-light images to some extent. However, the restoration of fine textures, especially in text regions [16], remains challenging, which is crucial for subsequent text detection [17] and recognition [18] tasks. As shown in Fig. 1, text detection is performed on both extremely low-light images and enhanced images using the TextBPN++ [19] and DBNet [20] methods as detectors. From Fig. 1(a), it can be observed that text detection on low-light images is often difficult to perform, whereas enhanced images enable text detection, as shown in Fig. 1(b).

However, there are additional difficulties in using paired data for image processing tasks such as dehazing, denoising, or low-light enhancement: 1) Simultaneously capturing damaged images and ground truth images of the same visual scene is extremely challenging, if not unrealistic (e.g., pairs of low-light and normal-light images). 2) Synthesizing low-light images required for normal image synthesis can be helpful, but such synthesized photos are often not realistic enough, leading to various artifacts when applying trained models to real-world low-light images. 3) Particularly for low-light enhancement, given a low-light image, there may not be a unique or well-defined normal-light image. For instance, any photo captured from dawn to dusk can be considered as a high-light version of a photo taken at midnight in the same scene. Considering these challenges, the overall objective of this paper is to enhance low-light photos with spatially varying lighting conditions and excessive (or insufficient) exposure, even in the absence of available paired training data.

Depending on the type of data, the problem of low-light image enhancement can be trained using paired or unpaired data. Some low-light image enhancement methods trained with paired data have shown promising results [21, 22]. However, the expensive data collection methods require greater ease of scalability. Currently, widely used low-light datasets include LOL [21]and SID [23], which are obtained by varying the camera's exposure time to capture both low-light and normal-light images. In practice, under weak lighting conditions and spatial variations, different exposure times still result in



(a) Input (b) Ours

Fig. 1 The DBNet text detection results before and after low-light image enhancement

deviations between natural low/normal-light images, accurate image mappings, and normal-light images. Clearly, this is a tedious and inefficient way of acquiring data. Thus, unpaired data image enhancement methods have attracted the attention of scholars [24–26]. However, these methods are prone to blurring, mosaic, and image distortion effects during image enhancement in low-light or even very dark scenes.

To meet the high-resolution input requirements of intelligent devices, object detection networks need to possess deeper or higher-dimensional feature maps. However, solely uploading and processing data in the cloud may result in unsatisfactory user experiences, especially considering the time sensitivity of mobile computing. Although current mobile phones have specific edge computing capabilities [27–29], they may struggle with complex computations in certain scenarios. To be specific, to handle highly dark scene situations, as many enhancement sub-networks as possible must be placed for dynamic enhancement. However, the number of sub-networks imposes a vast computational burden, which can only be performed in a cloud with sufficient computational resources. To address this issue, cloud computing is widely employed [30]. Thus, this paper proposes a deep learning method based on edge computing for text detection under low-light conditions, bridging the gap between the demands of object detection in low-light environments and the limited computational resources available on edge devices for scene understanding tasks. It helps to improve the visual perception of IoT devices in multiple fields, such as smart cities, autonomous driving, education and research. As illustrated in Fig. 2, a well-connected user establishes a wireless network or data transmission connection to the cloud server. The cloud server receives and processes user requests with the corresponding algorithmic calculations. In cases where multiple requests are received from multiple mobile users, the cloud server performs parallel computations in the cloud. Upon completion of the computations, real-time responses are provided to the users. The overall data training and testing framework maintains the accuracy of edge computing resources in text detection. Additionally, it transfers the time-consuming enhancement stage to the cloud, reducing time delays during the testing process.

Then inspired by EnlightenGAN [24] and investigated an end-to-end algorithm for low-light text image enhancement, called Unpaired text-attention Generative Adversarial Network (LAE-GAN). LAE-GAN consists of an attention-guided AGM-Net and ETAM as the generator, with global and local dual discriminators guiding the information. Firstly, this paper proposes a self-attention AGM-Net module based on M-Net [31], which extracts multi-scale features from the pre-enhanced image and attention map. Then, improvements are made to the sampling process, and a spatial attention map is incorporated to achieve noise suppression and brightness balancing. In addition, this paper introduces the Enhanced Text Attention Mechanism (ETAM). On one hand, the ETAM obtains a text attention map through text detection, aiming to constrain the text regions for clearer text information and improved overall image quality. On the other hand, the paper incorporates the Sobel edge detection operator to better restore the edge details of the text regions. The main contributions of this paper can be summarized as follows:

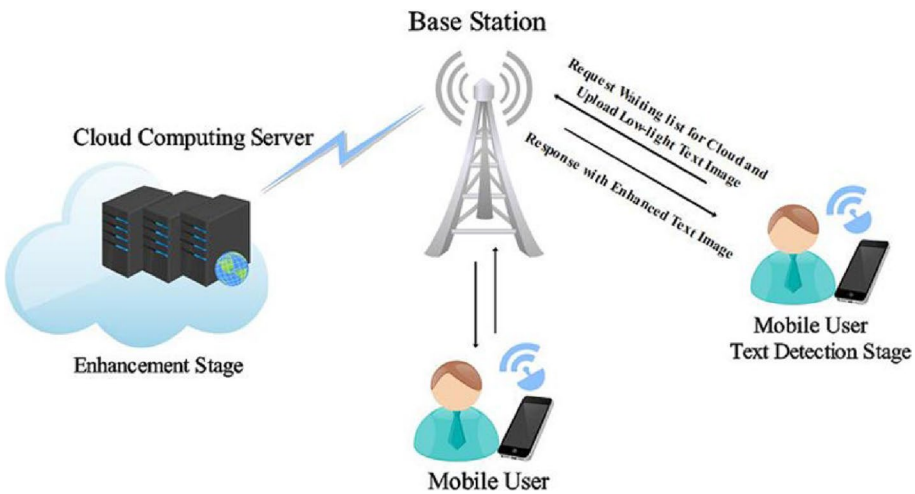


Fig. 2 The Cloud-Edge Computing Framework for Image Enhancement and Object Detection on Mobile Processing Devices

- We propose a deep learning method for text detection under low-light conditions, based on edge computing. This method maintains the accuracy of edge computing resources and target detection capabilities. Additionally, we transfer the time-consuming enhancement stage to the cloud, reducing the latency in testing.
- We introduce a non-paired text image enhancement method that addresses the difficulty of acquiring paired datasets while ensuring image quality.
- We introduce a novel AGM-Net network module to reduce noise in image enhancement and improve the accuracy of text detection in low-light scenes. Additionally, a Enhanced Text Attention Module (ETAM) based on non-paired images is proposed to enhance the brightness and details of text, thereby improving the accuracy of text detection.
- Through qualitative and quantitative experiments, the proposed method in this paper has been shown to outperform existing methods.

Related work

Low light image enhancement

Generally, low-light image enhancement can be divided into two main approaches: traditional methods and deep learning-based methods. In traditional methods, there are some classic approaches, such as utilizing histograms to represent the statistical information between pixels in an image and applying specific transformation operations to achieve uniform distribution of enhanced image pixels. Variants of these methods include AHE [22, 25], CLAHE [24], and others. To overcome their limitations, in the 1970s, Land [32] and others proposed the Retinex theory, which suggests that the perceived color of an object is related to its reflectance rather than the light projected onto the human eye. This theory implies that enhanced images should have color constancy. Building upon the Retinex theory, a series of improved algorithms have been developed, including single-scale Retinex (SSR), multi-scale Retinex (MSR), and color-restored multi-scale Retinex (MSRCR) [33].

Recently, significant improvements have been made in image enhancement using deep learning methods. Existing deep learning approaches mostly rely on training with paired datasets. Lore [34] proposed a multi-layer autoencoder (LL-net) that learns joint denoising and enhancement at the patch level under low-light conditions. Wei et al. [21] introduced an end-to-end framework that combines the Retinex theory with a network. HDR-Net [22] combines the idea of deep networks with paired bilateral grid processing and locally affine color transformations as supervision. In recent work, Zero-DCE net [25] enhances image details and reduces noise. However, Zero-DCE suffers from insufficient illumination enhancement for extremely low-light images, leading to visible noise artifacts.

Edge computing paradigm

The proposed cloud-edge computing framework, as illustrated in Fig. 2, aims to explore the feasibility of enhancing computing power at the network edge near the data source, leading to the emergence of various new computing models. These models can generally be categorized into three types: fog computing, cloudlets [35], and mobile edge computing (MEC), all of which share similar architectural concepts with edge computing. Stgyanarayanan et al. [36] effectively reduced service latency by leveraging large-scale IoT intelligent devices combined with edge computing architecture systems. In this regard, Sun et al. [37] proposed a novel method of mobile edge computing called edge IoT within the Internet of Things architecture, enabling efficient processing of massive mobile edge data streams. The low-latency characteristic of edge computing systems contributes to their wide application value in the field of video surveillance. Wang et al. [38] employed edge computing to collect and preprocess videos, utilizing distributed file systems for data storage and deep convolutional networks for data analysis, thereby effectively reducing system latency and improving image quality. Chen et al. [39] proposed a distributed surveillance system that combines deep learning with edge computing, effectively reducing network communication consumption and providing an efficient solution for surveillance video analysis.

Of particular interest to us is MEC, which serves as a beneficial complement to cloud infrastructures. MEC encourages computing operations to be conducted at the end of the IoT network and offers advantages such as long battery life, low transmission delay, low bandwidth consumption, and user privacy security. Based on these advantages, we have migrated and implemented the fault flow detection algorithm on edge devices, and experimental results demonstrate the superior performance of our framework [40–42].

The framework we propose enhances images by building servers in the cloud and performing detection algorithms at the terminal edge, aiming to meet the requirements of fast response and low bandwidth consumption.

Generative adversarial networks

In recent years, GANs have garnered significant attention in the field of low-light enhancement due to their successful applications in image synthesis and image translation tasks. Specifically, Pix2pix has demonstrated visually convincing results when provided with paired

training data in the target domain. However, acquiring such paired data in real-life scenarios poses significant challenges. Consequently, the introduction of cycle consistency loss in CycleGAN [43] has opened up possibilities for image-to-image translation without the need for paired data. To address the complexities of CycleGAN, Jiang et al. [24] proposed EnlightenGAN, which is the first unsupervised single-path generative adversarial network designed for low-light enhancement. This approach incorporates dual discriminators, namely the global discriminator and the local discriminator, along with self-regularized perceptual loss and an attention mechanism. EnlightenGAN aims to overcome the limitations associated with training without low-light/normal-light image pairs and the scarcity of paired datasets containing low-light and normal-light images. However, the algorithm is constrained by the inherent limitations of the generative adversarial network architecture, which may result in issues such as blurring and checkerboard artifacts when applied to high-resolution images.

Image denoising

The presence of noise in images significantly degrades their visual quality and has a notable impact on subsequent detection and recognition tasks. However, current denoising methods often rely on the addition of branches to the network architecture, which increases complexity and training parameters, resulting in a significant slowdown in training speed. A review [44] suggests that expanding the receptive field of the network can capture more contextual information, leading to improved denoising performance. Increasing the network's depth and width is a commonly employed technique to enlarge the receptive field, and dilated convolutions have proven effective in addressing this challenge. Moreover, the combination of convolutional neural networks (CNNs) with dimensionality reduction methods is widely employed for image denoising [45].

Algorithm for enhancing low-light text images

Currently, there are two commonly used enhancement methods for text images: spatial domain methods and frequency domain methods.

In spatial domain methods

the pixel values in the image are directly processed. Representative methods include grayscale transformation, histogram equalization, and smoothing techniques. Histogram equalization is a commonly used method that automatically enhances the contrast of an image by stretching the regions with higher grayscale densities, resulting in improved contrast for better visualization. However, this method has its drawbacks, especially for text grayscale images, as stretching the high grayscale value regions can disrupt the structure of the text and degrade the overall image quality. Researchers have proposed local histogram equalization methods to address this issue, but these methods often require traversing the entire image, leading to high time complexity and low efficiency. Kim et al. [46] introduced a histogram equalization method based on partially overlapping blocks, reducing the computational burden and improving the image quality. However, it may introduce block artifacts in the process.

In frequency domain methods

In essence, frequency domain methods treat the image as a two-dimensional signal and transform it to the frequency domain using Fourier Transform. By applying filters in the frequency domain, irrelevant signals are filtered out, while useful signals are preserved. The processed image is then transformed back to the spatial domain. Zhang et al. [47] proposed a novel fusion-based text enhancement method based on the understanding that color is sensitive to human perception, while frequency coefficients are sensitive to pixel-level variations. For each input image, an enhancement image is generated using the color space and another using the frequency domain. A new fusion method is then introduced to combine the two enhancement images, with weights determined based on text properties, resulting in the final fused image. However, due to limitations in automatic weight calculation, the proposed enhancement method may not perform optimally. Although the method effectively enhances text information by suppressing background details, it falls short in improving text detection performance. While the proposed method enhances text details, it is not sufficient for improving text detection performance.

Methodology

To address the inconvenience caused by paired datasets and the computational burden on edge and resource-constrained devices, we build a cloud server and perform detection algorithms on the edge to enhance images. Our approach mainly adopts the concept of Generative Adversarial Networks (GAN), where the generator enhances low-light images and the discriminator evaluates the generated images. Currently, deep learning methods have limited effectiveness in enhancing text regions. Therefore, we propose using a text attention mechanism to highlight text areas. Moreover, to address information loss caused by the U-Net architecture, we employ an improved M-Net network called AGM-Net. The overall structure consists of a generator composed of a Enhanced Text Attention Mechanism (ETAM) and

the improved M-Net network, and a discriminator composed of a global discriminator and local discriminator.

Cloud-edge computing framework

The computing framework structure based on the cloud server is shown in Fig. 3. This framework is designed for users with good network connections who establish a connection with the cloud server through wireless networks or data transmission. The cloud server performs algorithm calculations based on the requests submitted by the users. When the cloud server receives multiple requests from mobile users simultaneously, it conducts parallel computing in the cloud and provides real-time responses to the users after the calculations are completed.

In our research scenario, if a mobile user wants to determine the object categories in an image captured under poor lighting conditions, the proposed framework follows four steps: 1) Users locally submit image enhancement requirements to the cloud server and upload low-light images for further processing via wireless networks or data transmission; 2) The cloud server organizes proper parallel execution for multiple mobile users; 3) The cloud server sends a small-sized response to the user through the wireless network or mobile data packaged with data compression algorithms containing enhanced images processed by the cloud, thereby reducing the delay of network transmission, and sets the priority between the edge device and the cloud server during the transmission back to ensure that the data transmission and processing of the essential tasks are carried out at a higher priority. In contrast, the secondary tasks can be processed with low priority. When the computer server schedules computing workloads for multiple moving users, it is set to prioritise transmission for a particular user; 4) The mobile devices, acting as edge devices, use customized correlation algorithms to detect the images and provide accurate detection results to the users.

To cope with poor lighting conditions, a robust performance-enhancing algorithm is required for image enhancement. However, the computational burden of a large number of images can only be handled in the cloud with sufficient computing resources. In practice, there can be significant disparities in the computing power of mobile devices, leading us to design different versions of algorithms that can run locally for fast response.

Network architecture

In this section, we propose a novel end-to-end framework for enhancing extremely low-light text images, enabling simultaneous enhancement of both the text and the overall image under challenging lighting conditions. The network architecture is illustrated in Fig. 4. Our framework consists of a Enhanced Text Attention Module (ETAM) and the AGM-Net generator, complemented by a global discriminator and a local discriminator forming the discriminator structure. For the generator, we introduce a method that incorporates a text attention map obtained through a generic text detection approach. Additionally, we employ the Sobel operator loss between the enhanced intermediate images obtained using Zero-DCE and the final generated images to highlight text details. Furthermore, we utilize the attention-guided AGM-Net as the generator, which enriches contextual information from different resolution images, mitigating

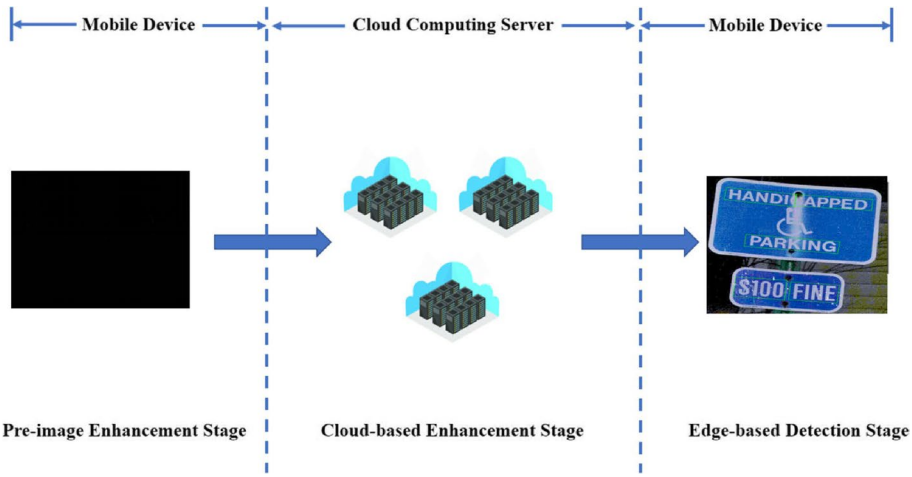


Fig. 3 Low-light Text image Enhancement Generative Adversarial Network for Cloud Computing, the user sends an image to the cloud server through the mobile phone, and the cloud server returns the algorithmically enhanced picture to the mobile terminal, and the mobile terminal returns the detection result to the user terminal

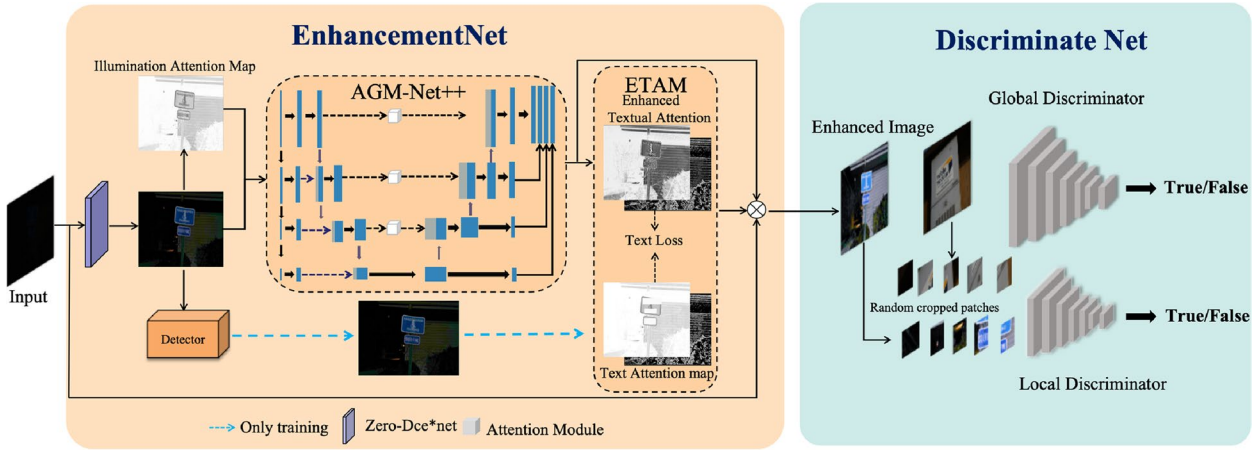


Fig. 4 Network overview. Zero-DCE[a] [25] is retrained Zero-DCE net. The dashed line is passed through the text detector for Zero-DCE enhanced text as a text attention map. The AGM-Net module is our image enhancement and denoising module

spatial losses caused by sampling and effectively addressing noise issues.

Attention-guided AGM-Net

In extremely dark environments, images suffer from low visibility, making it difficult to obtain useful information visually. Corresponding clear images are challenging to obtain in real-world scenarios. U-Net [48] is widely used as the foundational infrastructure for low-light image enhancement. However, U-Net introduces significant noise and severe spatial information loss while recovering illumination, leading to substantial instances of missed detections in subsequent text detection tasks. Therefore, this paper proposes a novel attention-guided AGM network module to address the noise and information loss issues introduced by the U-Net architecture. The network structure of the generator is illustrated in Fig. 5. It can be considered as an improved hierarchical model architecture of U-Net. The module achieves good denoising effects on images. AGM-Net introduces additional gatepost feature pathways in both the encoder and decoder to enrich contextual information at different image resolutions and improve spatial loss caused by sampling. Furthermore, the aim is to enhance dark areas rather than bright areas, ensuring that the output image is neither overexposed nor underexposed. In this paper, the luminance channel L of the input RGB image is extracted and normalized to [0, 1]. The self-regularized attention map is obtained by 1-L (element-wise difference). The attention

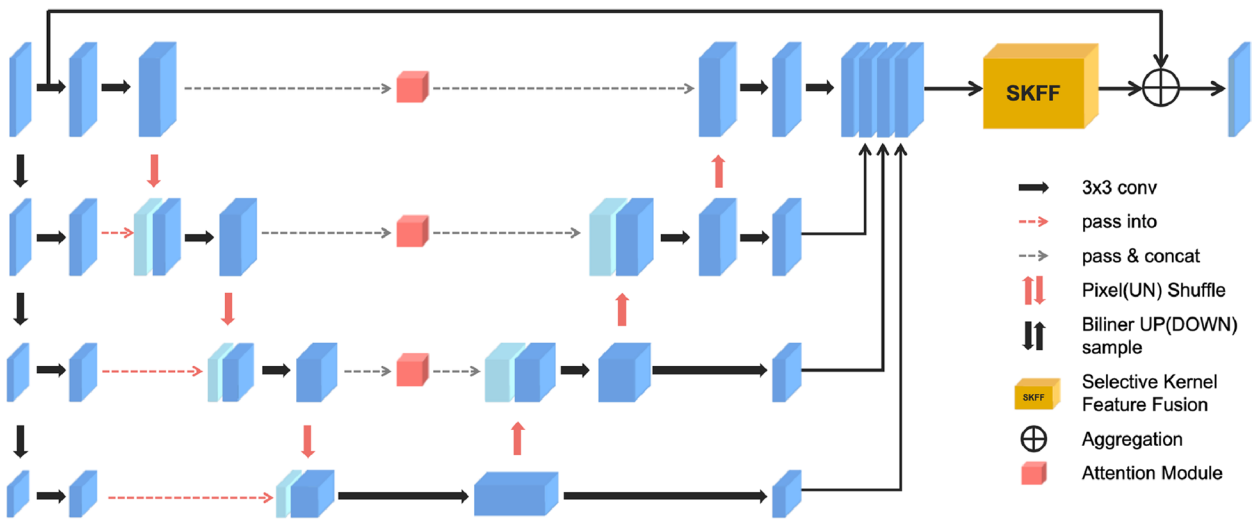


Fig. 5 AGM-Net. AGM-Net is based on a self-regularized attention-guided image enhancement module, and we have four layers in the proposed AGM-Net

maps are resized to match each feature map and multiplied with all intermediate feature maps and the output image. In U-Net, max pooling is utilized for downsampling. However, as the restoration task involves pixel-level visual processing, downsampling is not suitable as it leads to significant loss of spatial information. Therefore, AGM-Net employs pixel UNshuffle downsampling and bilinear downsampling in the gatepost pathway, which introduces more diversity to the cascaded features. Moreover, the design of gatepost feature pathways in AGM-Net helps mitigate spatial information loss. Furthermore, by incorporating self-regularized attention maps in the skip connections, this study achieves a balance between insufficient enhancement in input image increments and overexposed regions during the enhancement process.

Text attention mechanism

Acquiring paired datasets for extremely dark conditions is challenging and requires significant annotation costs. Therefore, this study introduces a text attention mechanism and an edge detection map as the text attention map to highlight text details.

Text attention mechanism

In this study, the proposed method utilizes the text detection algorithm to obtain text annotation information, which serves as the text attention map. The generator takes low-light images under extreme conditions as input and reconstructs the incremental image. The input and output image sizes are set to 600∗400 pixels. Specifically, the input image is preprocessed using Zero-DCE[a] for initial illumination enhancement. Then, a text detection algorithm is applied to the preliminarily enhanced image to obtain text annotation information. In this study, the pixel values of the identified text regions in the RGB image are set to 200, and the illumination channel I is normalized to the range [0,1]. and then 1-I (element difference) is used as the text attention map I'. The enhanced image obtained through AGM-net++ is also processed by normalizing the illumination channel L to the range [0,1]. The 1-I (element difference) and I is used as the self-regularized attention map L'. By constraining the text attention map and the self-regularized attention map, the training network can focus more on the text regions.To further emphasize the text regions, this study also employs L1 loss between I' and L' as follows:

$$\mathcal{L}_{TEXT} = \| I' - L' \| \tag{1}$$

Edge attention map

To address the impact of upsampling and downsampling on image details and to better highlight the text information for subsequent detection and recognition tasks, this paper introduces the Sobel edge detection operator into the network to extract the edge information of the text. The Sobel operator is a linear filter that is easy to implement and computationally efficient, making it easier to perform large-scale image processing tasks on cloud servers, reducing the processing cost of the algorithm while effectively highlighting local features in the image, such as text outlines that typically appear at localised locations in the image. Specifically, the Sobel operator is applied to extract the edge information map E' from the intermediate enhanced image obtained through Zero-DCE. Similarly, the Sobel operator is applied to extract the edge map O' from the final generated image. By constraining the distance between E' and O', this paper ensures the preservation of text region details. The constraint is implemented using an $L2$ loss as follows:

$$\mathcal{L}_{EDGE} = \| E' - O' \| \tag{2}$$

Loss function

In previous studies on enhancing dark scene images, it has been observed that overexposure and underexposure often occur in low-light background regions. This finding suggests that relying solely on a global discriminator may not adequately capture the desired image fitting. Inspired by EnlightenGAN [24], this paper proposes the adoption of a global-local discriminator structure. Specifically, for the local discriminator, five random regions are cropped and a relativistic discriminator [49] is employed. This discriminator estimates the probability that real data is more realistic than fake data, thereby indicating the generator's ability to synthesize fake images that are more realistic than real images. Additionally, LSGAN [50] is utilized as the adversarial loss in this work. Finally, the loss functions for the global discriminator D and the generator G are defined as follows:

$$\begin{aligned}\mathcal{L}_D^{\text{Global}} &= E_{x_r\sim P}\left[\left(\sigma\left(C(x_r) - \mathbb{E}_{x_f\sim\mathbb{Q}}\left[C\left(x_f\right)\right]\right) - 1\right)^2\right] \\ &+ \mathbb{E}_{x_f\sim\mathbb{Q}}\left[\left(\sigma\left(C\left(x_f\right) - \mathbb{E}_{x_r\sim\mathbb{P}}[C(x_r)]\right) - 0\right)^2\right],\end{aligned} \tag{3}$$

$$\begin{aligned}\mathcal{L}_G^{\text{Global}} &= E_{x_f\sim Q}\left[\left(\sigma\left(C\left(x_f\right) - \mathbb{E}_{x_r\sim\mathbb{P}}[C(x_r)]\right) - 1\right)^2\right] \\ &+ \mathbb{E}_{x_r\sim\mathbb{P}}\left[\left(\sigma\left(C(x_r) - \mathbb{E}_{x_f\sim\mathbb{Q}}\left[C\left(x_f\right)\right]\right) - 0\right)^2\right],\end{aligned} \tag{4}$$

Where $C(\cdot)$ denotes the discriminator network, x_r and x_f are sampled from the true-false distribution, where $\mathbb{E}_{x_f\sim\mathbb{Q}}$ denotes the generated data probability $\mathbb{E}_{x_r\sim\mathbb{P}}$ denotes the true data probability.

In the unpaired setting of this paper, the losses for both the global and local discriminators are guided by feature preservation to constrain the VGG feature distance between low-light and its enhanced normal-light

counterparts. As a result, the overall loss function used to train LAE-GAN is therefore formulated as follows. Hence, the overall loss function for training this network is defined as:

$$L_{total} = \mathcal{L}_{TEXT} + \mathcal{L}_{EDGE} + \mathcal{L}_{SFP}^{\text{Global}} + \mathcal{L}_{SFP}^{\text{Local}} + \mathcal{L}_{G}^{\text{Global}} + \mathcal{L}_{G}^{\text{Local}} \quad (5)$$

Experiments

Our method is based on the pix2pix model to enhance low-light text images, combining self-attention mechanisms and gradient loss in the network to make it more focused on the recovery of text edge information. This section will demonstrate the effectiveness of the different modules and the validity of the proposed method experimentally.

Datasets and training setting

Due to the network's ability to train on non-paired low/normal-light images, this paper was able to gather a larger amount of non-paired training datasets. The LOL and SID datasets were used for training, where LOL represents images captured under low-light conditions and SID represents images captured under extremely dark conditions. The low-light images were captured using short exposure times, while the normal-light images were collected using long exposure times. Specifically, this paper curated a non-paired dataset consisting of 300 low-light images and 300 normal-light images. The entire network in this paper was trained using a single GTX 3090 GPU. The Adam optimizer was utilized, starting with a learning rate of 1e-4 for the initial 100 epochs. Subsequently, another 100 epochs were conducted with the learning rate decayed to 0. In terms of evaluation metrics, this paper focused on peak signal-to-noise ratio (PSNR), structural similarity index (SSIM), and no-reference image quality assessment (NIQE) for quantitative comparisons. Higher PSNR and SSIM values indicate superior performance, while a lower NIQE value indicates higher image quality.

Visual comparison

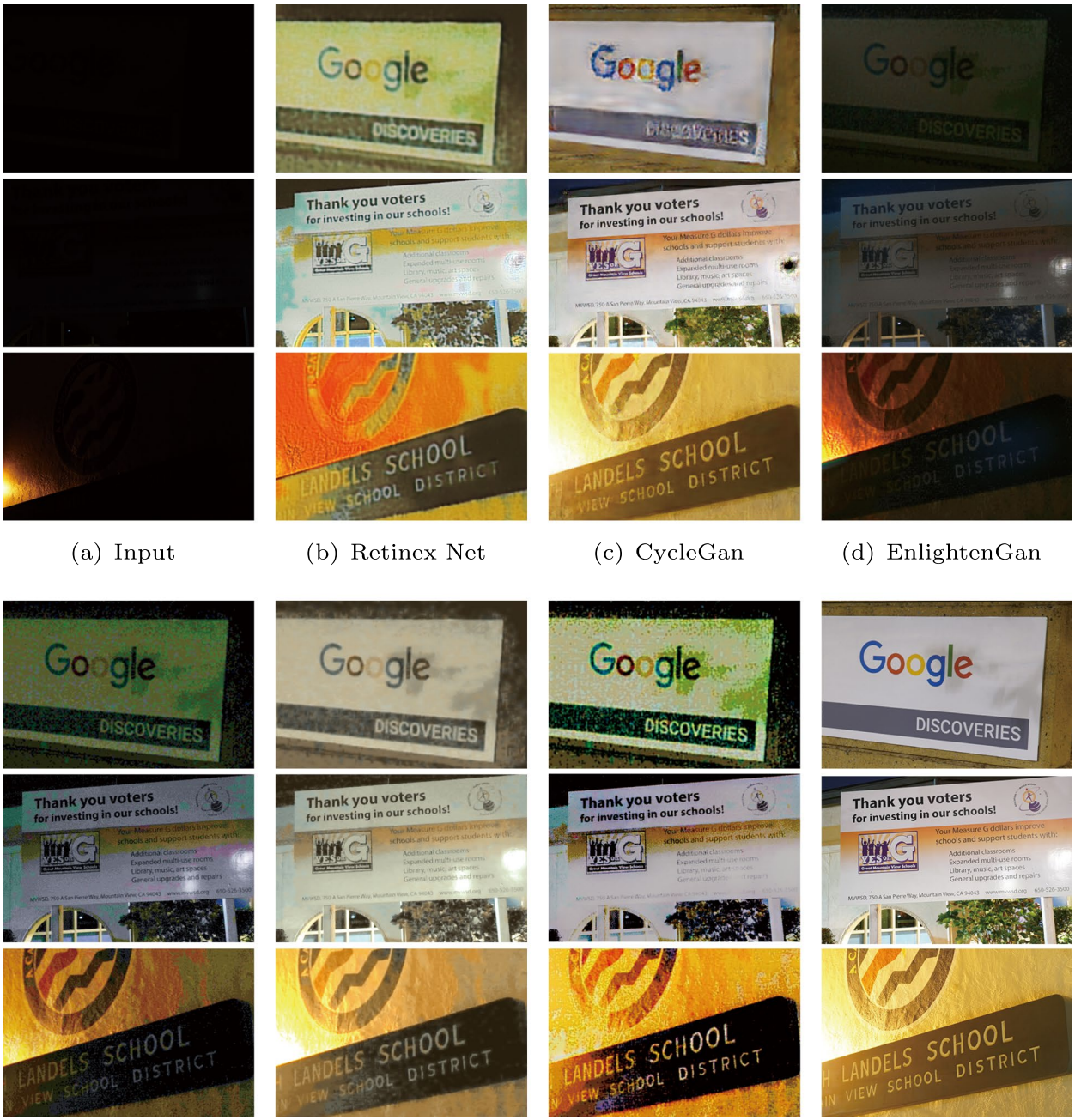
To evaluate the effectiveness of the proposed LAE-GAN method, a comparison was conducted with existing low-light enhancement methods on SID datasets, as shown in Fig. 6. The first column represents the original input image, while the second to sixth columns display the enhanced images obtained using the pre-trained Retinex [33], Cycle-GAN [43], EnlightenGAN [24], Zero-DCE [25], IAT [26], and LAE-GAN, respectively. Additionally, the last column showcases the reference normal-light image. From the figure, it can be observed that Cycle-GAN and IAT produce enhanced images with severe distortion. EnlightenGAN and Zero-DCE are capable of handling low-light images and achieving high-quality results. However, these methods require more lighting enhancement and introduce significant noise when dealing with low-lit images. In comparison, the proposed LAE-GAN demonstrates overall improvement in image quality, excels in restoring text details, and handles images with varying levels of lighting intensity and noise better than other methods. Additionally, this paper evaluates the quality metrics of the enhanced images obtained by each model, as shown in Table 1. Our proposed LAE-GAN achieves the best results in terms of SSIM and NIQE metrics and performs competitively in terms of PSNR. These results highlight the focus of our model on restoring text details.

Ablation experiments

To demonstrate the effectiveness of each component in LAE-GAN, We conducted ablation studies. Specifically, the text attention mechanism and AGM-Net were individually removed, and the corresponding image quality metrics were evaluated. The results, presented in Table 2, reveal that both the proposed AGM-Net module and ETAM module significantly improved the overall image quality, The combination of the two modules significantly increased PSNR and SSIM by 1.85 and 0.23, respectively. In contrast, other methods require more lighting enhancement and introduce significant noise when handling low-lit images. Comparatively, the proposed LAE-GAN method in this paper enhances the overall image quality, excels in restoring text details, and performs better in handling images with varying lighting intensities and noise levels. Furthermore, the quality metrics of the enhanced images obtained by each model were tested, as shown in Table 1. The results indicate that our proposed UT-GAN performs comparably to the state-of-the-art methods on the extremely low-light dataset SID. Although the PSNR metric does not achieve the highest numerical value, the SSIM metric yields the best result with a 0.12 improvement over the IAT method.

Comparison of scene text detection

To assess the effectiveness of LAE-GAN in text detection, we employed the DBNet [20] and TextBPN++ [19] text detection algorithms for validation, as depicted in Fig. 7. The first and second rows display the text detection results obtained using DBNet and TextBPN++ , respectively. Fig. 7 (a) represents the original low-light image, while Fig. 7 (b) and (c) illustrate the detection results of the enhanced images using Zero-DCE and IAT methods. Fig. 7 (d) showcases the detection results of the enhanced images using the proposed LAE-GAN method. It can be observed that Zero-DCE yield unsatisfactory



(a) Input (b) Retinex Net (c) CycleGan (d) EnlightenGan

(e) Zero-DCE (f) IAT (g) Ours (h) GT

Fig. 6 Visual comparison of existing methods on the SID dataset, revealing that Retinex, Cycle-GAN and IAT introduce distortion in the enhanced images. Zero-DCE and EnlightenGAN show suboptimal enhancement results with noticeable noise artifacts. On the other hand, LAE-GAN excels in noise reduction and texture detail restoration, delivering the most favorable outcome in terms of image quality

enhancement of the text regions in the image, introducing significant noise and resulting in poor text detection performance. Conversely, LAE-GAN demonstrates impressive results in terms of overall brightness and text detail restoration. Quantitative metrics are presented in Table 3, employing precision, recall, and F1-score (Hmean) as evaluation criteria. LAE-GAN enhances the text structure better, improving the accuracy and reaching the optimum, but the recall value is less favourable. The specific reasons for this will be analysed later in the Discussion section.

Table 1 Performance evaluation of different methods on SID and LOL dataset. [a]means that the model is pre-trained. We highlight the best and second best results using red and blue text, respectively

Method	SID		LOL	
	PSNR	SSIM	PSNR	SSIM
LIME(2017)	14.92	0.42	15.32	0.46
Retinex Net[a](2018)	15.43	0.38	15.69	0.90
Cyclegan[a](2018)	15.51	0.46	18.73	0.92
EnlightenGAN[a](2021)	14.62	0.41	17.74	0.68
Zero-DCE[a](2020)	15.43	0.47	16.42	0.54
IAT[a](2022)	16.73	0.53	22.80	0.97
Ours	16.47	0.65	18.54	0.73

Table 2 Ablation study on SID. ETAM and AGM are abbreviations for textual attention mechanism and self-attentive guided AGM-net, respectively

Method	SSIM	PSNR	NIQE
Baseline	14.62	0.42	11.34
ETAM+Baseline	14.72	0.43	11.31
AGM-Net+Baseline	16.21	0.57	10.74
Baseline+ETAM+AGM-Net	16.47	0.65	10.21

Complexity analysis

To verify the efficiency of the proposed method, we select 50 high-definition text images of very dark scenes with a resolution of 1616x1080, where their sizes are in the range of 0.5 to 1.5 M. Then the images are compressed and transmitted within about 10 to 15 seconds over a 4G cellular network, and within about 3 to 5 seconds when using the cloud. For the enhancement phase, on a server equipped with a NVIDIA GTX 3090, the average time spent per image is 0.15 seconds. It only uses local and consumes an average of 5.12 seconds. It can be seen that the cloud server-based model is faster than processing images locally, proving that the proposed method effectively reduces the computational burden on mobile devices.

Disscussion

Currently, mobile devices with limited resources face challenging text image enhancement techniques. In this study, to enable mobile devices to perform more efficiently and accurately in subsequent text detection and recognition tasks, such as night-time road sign detection in autonomous driving, in the face of the challenges of extremely dim environments, we propose the LAE-GAN model, which is deployed on a cloud server for efficient parallel computing.

However, as shown in Tables 1, 2, and 3, experiments also found differences with other methods. Firstly, for the enhancement phase, the LAE-GAN model focuses more on text image recovery in very dark scenes, and facing the LOL text-free dataset, some may not be balanced enough for light recovery, or the image is not smooth enough. Secondly, the model enhances the edges of the overall image structure at the same time when performing text structure recovery, resulting in text images that are not smooth enough, so the PSNR is lower compared to other models, but the SSIM value reaches the optimum.

Then, for the detection phase, after a lot of exploration, we have come up with the possible reasons for the poor

(a) Input (b) Zero-DCE[a] (c) IAT[a] (d) Ours

Fig. 7 Vision comparison of text detection algorithms. on different low-light image enhancement methods. [a]represents that the model was pre-trained in the enhancement phase

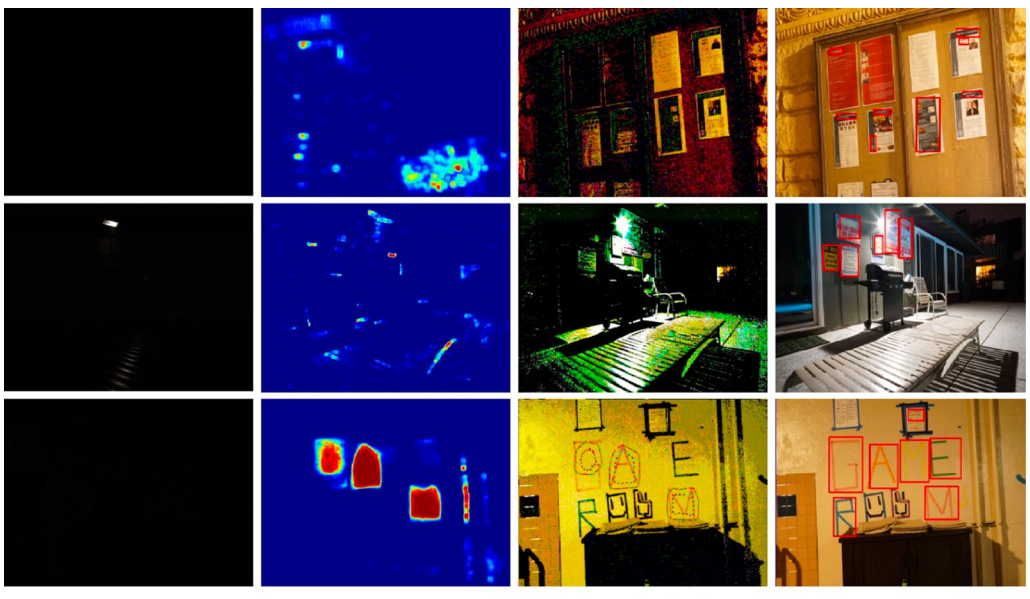
Table 3 Performance evaluation of different methods on the SID dataset. [a]represents that the model was pre-trained in the enhancement phase. We highlight the best and second best results using red and blue text, respectively

Dection Methods	Enhanced Methods	precision	recall	F1score
DBNet(2019)	Cyclegan[a] (2018)	0.54	0.31	0.39
	Zero-DCE[a](2020)	0.48	0.23	0.31
	EnlightenGAN[a](2021)	0.58	0.25	0.35
	IAT[a](2022)	0.57	0.29	0.38
	Ours	0.60	0.22	0.32
TextBPN++(2021)	Cyclegan[a] (2018)	0.45	0.27	0.34
	Zero-DCE[a](2020)	0.33	0.21	0.26
	EnlightenGAN[a](2021)	0.44	0.32	0.37
	IAT[a](2022)	0.45	0.35	0.39
	Ours	0.50	0.15	0.23

recall values as follows: 1) When labelling text images, some of the texts in the labels are too small or dense, and in the enhanced model generator, we used the universal detection model (EAST) to obtain the text attention map due to the poor processing of the EAST model in the face of complex text structures such as curved text, which may result in the structure of the small text is not clear so that the current detection methods can not be such an accurate detection. 2) LAE-GAN enhances the text structure simultaneously as it enhances the image details, resulting in a low level of differentiation between small text areas and the background in the image, which may affect the judgement of the detection model. The specific case is shown in Fig. 8. We plan to enhance the attention to small text structures by optimizing the detection block and noise reduction capability of AGMNet in future work.

Conclusion and future work

In this paper, a novel text image enhancement network called LAE-GAN is proposed for extremely dark scene conditions. This method demonstrates excellent generalization and operability without the need for paired training data. The paper introduces self-feature preservation loss, which constrains the feature distance between the low-light input and its enhanced normal-light output by applying the loss to the generated images from the generator. This approach enables training with non-paired datasets, addressing challenges in data collection, handling inaccuracies in high-exposure images compared to normal-light images, and reducing the waste of manual resources required for annotation. The paper constructs the Enhanced Text Attention Mechanism (ETAM), which utilizes the intermediate image from Zero-DCE for text detection. This mechanism obtains a text attention map to constrain the generator's generated images, highlighting text regions and recovering the text image's details better utilizing the Sobel edge detection operator. Furthermore, we introduce a novel AGMNet network module with a branch-free structure that achieves outstanding enhancement and denoising performance. The overall image quality is significantly improved by utilising gate pathways and self-attention mechanisms.



(a) Origin (b) Attention Map (c) Ours (d) GT

Fig. 8 Failure cases of minor text image detection

The experimental results show that the method in this paper has made significant progress in text enhancement, especially in handling text images in extremely dark scenes. However, the recovery of text structure is challenging, and the recovery process enhances the edges of the image structure simultaneously, which may cause parts of the image to be insufficiently smooth and, accordingly, cause unsatisfactory text detection. In the future, conducting in-depth research based on improved text structure reconstruction, multimodal information fusion, image evaluation metrics, and cloud computing frameworks is worthwhile. To better apply the effect of text detection in different scenarios, we also plan to construct a larger dataset with more diverse coverage. These research directions will help to improve the performance of model and extend its application in different fields.

Acknowledgements
This work is supported by the Scientific Research Foundation of Chongqing University of Technology, Chongqing Postgraduate Innovation Fund. Chongqing Technology Innovation and Application Development under Grant, Natural Science Foundation of Chongqing, China under Grant and Chongqing Postgraduate Innovation Fund. I would like to thank the reviewers and editors for their guidance and help in the revision process.

Authors' contributions
X. M. wrote the main manuscript text; H. Y. Modification prepared visualization and helped perform the experiment; X. P. performed the experiment of text detection; H. Z. performed the experiment of low-light enhancement; F.X. Modification. All authors reviewed the manuscript.

Funding
This work is supported by the Scientific Research Foundation of Chongqing University of Technology, 2022ZDZ012, Chongqing Technology Innovation and Application Development under Grant No.cstc2021jscx-dxwtBX0018, Natural Science Foundation of Chongqing, China under Grant No.CSTB2022NSCQ-MSX0493 and Chongqing Postgraduate Innovation Fund, CYS23677.

Availability of data and materials
All data generated or analysed during this study are included in this published article.

Declarations

Ethics approval and consent to participate
Not applicable.

Competing interests
The authors declare no competing interests.

Received: 28 July 2023 Accepted: 22 October 2023
Published online: 18 November 2023

References

1. Sandhu AK (2021) Big data with cloud computing: Discussions and challenges. Big Data Min Analytics 5(1):32–40
2. Mousavi SN, Chen F, Abbasi M, Khosravi MR, Rafiee M (2022) Efficient pipelined flow classification for intelligent data processing in iot. Digit Commun Netw 8(4):561–575
3. Song W, Wu Y, Cui Y, Liu Q, Shen Y, Qiu Z, Yao J, Peng Z (2022) Public integrity verification for data sharing in cloud with asynchronous revocation. Digit Commun Netw 8(1):33–43
4. Liu Y, Wu H, Rezaee K, Khosravi MR, Khalaf OI, Khan AA, Ramesh D, Qi L (2022) Interaction-enhanced and time-aware graph convolutional network for successive point-of-interest recommendation in traveling enterprises. IEEE Trans Ind Inform 19(1):635–643
5. Qi L, Liu Y, Zhang Y, Xu X, Bilal M, Song H (2022) Privacy-aware point-of-interest category recommendation in internet of things. IEEE Internet Things J 9(21):21398–21408
6. Liu Y, Zhou X, Kou H et al (2023) Privacy-Preserving Point-of-Interest Recommendation based on Simplified Graph Convolutional Network for Geological Traveling[J]. ACM Transactions on Intelligent Systems and Technology
7. Xue M, Huang Z, Liu RZ, Lu T (2021) A Novel Attention Enhanced Residual-In-Residual Dense Network for Text Image Super-Resolution. 2021 IEEE International Conference on Multimedia and Expo (ICME), Shenzhen, p. 1–6. https://doi.org/10.1109/ICME51207.2021.9428128
8. Xue M, Shivakumara P, Zhang C, Lu T, Pal U (2019) Curved text detection in blurred/non-blurred video/scene images. Multimed Tools Appl 78:25629–25653
9. Chen Y, Zhao F, Lu Y, Chen X (2022) Dynamic task offloading for mobile edge computing with hybrid energy supply. Tsinghua Sci Technol 28(3):421–432
10. Chen Y, Xing H, Ma Z, Chen X, Huang J (2022) Cost-efficient edge caching for noma-enabled iot services. China Commun
11. Zhu E, Zhang J, Yan J, Chen K, Gao C (2022) N-gram malgan: Evading machine learning detection via feature n-gram. Digit Commun Netw 8(4):485–491
12. Zhang S, Yao L, Sun A, Tay Y (2019) Deep learning based recommender system: A survey and new perspectives. ACM Comput Surv (CSUR) 52(1):1–38
13. Kim J, Lee JK, Lee KM (2016) Accurate Image Super-Resolution Using Very Deep Convolutional Networks. Proceedings of the Ieee Computer Society Conference on Computer Vision and Pattern Recognition V2016-december. p 1646–1654. https://doi.org/10.1109/CVPR.2016.182
14. Zhang K, Zuo W, Chen Y, Meng D, Zhang L (2017) Beyond a gaussian denoiser: Residual learning of deep cnn for image denoising. IEEE Trans Image Process 26(7):3142–3155
15. Tao X, Gao H, Shen X, Wang J, Jia J (2018) Scale-recurrent network for deep image deblurring. In: Proceedings of the IEEE conference on computer vision and pattern recognition. pp 8174–8182
16. Hsu PH, Lin CT, Ng CC, Kew JL, Tan MY, Lai SH, Chan CS, Zach C (2022) Extremely low-light image enhancement with scene text restoration. In: 2022 26th International Conference on Pattern Recognition (ICPR). IEEE, pp 317–323
17. Wang W, Xie E, Song X, Zang Y, Wang W, Lu T, Yu G, Shen C (2019) Efficient and accurate arbitrary-shaped text detection with pixel aggregation network. In: Proceedings of the IEEE/CVF international conference on computer vision. pp 8440–8449
18. Baek J, Kim G, Lee J, Park S, Han D, Yun S, Oh SJ, Lee H (2019) What is wrong with scene text recognition model comparisons? dataset and model analysis. In: Proceedings of the IEEE/CVF international conference on computer vision. pp 4715–4723
19. Zhang SX, Zhu X, Yang C, Wang H, Yin XC (2021) Adaptive boundary proposal network for arbitrary shape text detection. In: Proceedings of the IEEE/CVF International Conference on Computer Vision. pp 1305–1314
20. Liao M, Wan Z, Yao C, Chen K, Bai X (2020) Real-time scene text detection with differentiable binarization. In: Proceedings of the AAAI conference on artificial intelligence, vol 34. pp 11474–11481
21. Wei C, Wang W, Yang W, Liu J (2018) Deep retinex decomposition for low-light enhancement. arXiv preprint arXiv:1808.04560
22. Gharbi M, Chen J, Barron JT, Hasinoff SW, Durand F (2017) Deep bilateral learning for real-time image enhancement. ACM Trans Graph (TOG) 36(4):1–12
23. Chen C, Chen Q, Xu J, et al (2018) Learning to see in the dark[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 3291-3300.
24. Jiang Y, Gong X, Liu D, Cheng Y, Fang C, Shen X, Yang J, Zhou P, Wang Z (2021) Enlightengan: Deep light enhancement without paired supervision. IEEE Trans Image Process 30:2340–2349

25. Guo C, Li C, Guo J, Loy CC, Hou J, Kwong S, Cong R (2020) Zero-reference deep curve estimation for low-light image enhancement. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. pp 1780–1789
26. Cui Z, Li K, Gu L, Su S, Gao P, Jiang Z, Qiao Y, Harada T (2022) Illumination adaptive transformer. arXiv preprint arXiv:2205.14871
27. Bi R, Liu Q, Ren J, Tan G (2020) Utility aware offloading for mobile-edge computing. Tsinghua Sci Technol 26(2):239–250
28. Huang J, Lv B, Wu Y, Chen Y, Shen X (2021) Dynamic admission control and resource allocation for mobile edge computing enabled small cell network. IEEE Trans Veh Technol 71(2):1964–1973
29. Qi L, Lin W, Zhang X, Dou W, Xu X, Chen J (2022) A correlation graph based approach for personalized and compatible web apis recommendation in mobile app development. IEEE Trans Knowl Data Eng
30. Xu Z, Zhu D, Chen J, Yu B (2022) Splitting and placement of data-intensive applications with machine learning for power system in cloud computing. Digit Commun Netw 8(4):476–484
31. Mehta R, Sivaswamy J (2017) M-net: A convolutional neural network for deep brain structure segmentation. In: 2017 IEEE 14th international symposium on biomedical imaging (ISBI 2017). IEEE, pp 437–440
32. Land EH, McCann JJ (1971) Lightness and retinex theory. Josa 61(1):1–11
33. Li C, Guo C, Han L, Jiang J, Cheng MM, Gu J, Loy CC (2021) Low-light image and video enhancement using deep learning: A survey. IEEE Trans Pattern Anal Mach Intell 44(12):9396–9416
34. Lore KG, Akintayo A, Sarkar S (2017) LLnet: A deep autoencoder approach to natural low-light image enhancement. Pattern Recogn 61:650–662
35. Laghari AA, Jumani AK, Laghari RA (2021) Review and state of art of fog computing. Arch Comput Methods Eng 28(5):3631–3643
36. Satyanarayanan M (2017) The emergence of edge computing. Computer 50(1):30–39
37. Sun X, Ansari N (2016) Edgeiot: Mobile edge computing for the internet of things. IEEE Commun Mag 54(12):22–29
38. Wang R, Tsai WT, He J, Liu C, Li Q, Deng E (2019) A video surveillance system based on permissioned blockchains and edge computing. In: 2019 IEEE international conference on big data and smart computing (BigComp). IEEE, pp 1–6
39. Chen J, Li K, Deng Q, Li K, Philip SY (2019) Distributed deep learning model for intelligent video surveillance systems with edge computing. IEEE Trans Ind Inform
40. Chen C, Liu B, Wan S, Qiao P, Pei Q (2020) An edge traffic flow detection scheme based on deep learning in an intelligent transportation system. IEEE Trans Intell Transp Syst 22(3):1840–1852
41. Wan S, Ding S, Chen C (2022) Edge computing enabled video segmentation for real-time traffic monitoring in internet of vehicles. Pattern Recogn 121:108146
42. Chen C, Liu L, Wan S, Hui X, Pei Q (2021) Data dissemination for industry 4.0 applications in internet of vehicles based on short-term traffic prediction. ACM Trans Internet Technol (TOIT) 22(1):1–18
43. Zhu JY, Park T, Isola P, Efros AA (2017) Unpaired image-to-image translation using cycle-consistent adversarial networks. In: Proceedings of the IEEE international conference on computer vision. pp 2223–2232
44. Wang T, Sun M, Hu K (2017) Dilated deep residual network for image denoising. In: 2017 IEEE 29th international conference on tools with artificial intelligence (ICTAI). IEEE, pp 1272–1279
45. Yuan Q, Zhang Q, Li J, Shen H, Zhang L (2018) Hyperspectral image denoising employing a spatial-spectral deep residual convolutional neural network. IEEE Trans Geosci Remote Sens 57(2):1205–1218
46. Kim JY, Kim LS, Hwang SH (2001) An advanced contrast enhancement using partially overlapped sub-block histogram equalization. IEEE Trans Circ Syst Video Technol 11(4):475–484
47. Zhang C, Shivakumara P, Xue M, Zhu L, Lu T, Pal U (2018) New fusion based enhancement for text detection in night video footage. In: Advances in Multimedia Information Processing–PCM 2018: 19th Pacific-Rim Conference on Multimedia, Hefei, China, September 21-22, 2018, Proceedings, Part III 19. Springer, pp 46–56
48. Ronneberger O, Fischer P, Brox T (2015) U-net: Convolutional networks for biomedical image segmentation. In: Medical Image Computing and Computer-Assisted Intervention–MICCAI 2015: 18th International Conference, Munich, Germany, October 5-9, 2015, Proceedings, Part III 18. Springer, pp 234–241
49. Jolicoeur-Martineau A (2018) The relativistic discriminator: a key element missing from standard gan. arXiv preprint arXiv:1807.00734
50. Mao X, Li Q, Xie H, Lau RY, Wang Z, Paul Smolley S (2017) Least squares generative adversarial networks. In: Proceedings of the IEEE international conference on computer vision. pp 2794–2802

Publisher's Note