

Article

Improved Lightweight Zero-Reference Deep Curve Estimation Low-Light Enhancement Algorithm for Night-Time Cow Detection

Zijia Yu ¹, Yangyang Guo ^{2,3}, Liyuan Zhang ^{4,*}, Yi Ding ², Gan Zhang ² and Dongyan Zhang ^{2,5,*} ¹ School of Information Engineering, Suzhou University, Suzhou 234000, China² School of Internet, Anhui University, Hefei 230039, China; guoyangyang113529@ahu.edu.cn (Y.G.)³ Fin China-Anhui University Joint Laboratory for Financial Big Data Research, Hefei 230039, China⁴ School of Agricultural Engineering, Jiangsu University, Zhenjiang 212013, China⁵ College of Mechanical and Electronic Engineering, Northwest A&F University, Yangling 712100, China

* Correspondence: liyuan.zhang@ujs.edu.cn (L.Z.); zhangdy@ahu.edu.cn (D.Z.)

Abstract: With the advancement of agricultural intelligence, dairy-cow farming has become a significant industry, and the application of computer vision technology in the automated monitoring of dairy cows has also attracted much attention. However, most of the images in the conventional detection dataset are high-quality images under normal lighting, which makes object detection very challenging in low-light environments at night. Therefore, this study proposed a night-time detection framework for cows based on an improved lightweight Zero-DCE (Zero-Reference Deep Curve Estimation) image enhancement network for low-light images. Firstly, the original feature extraction network of Zero-DCE was redesigned with an upsampling structure to reduce the influence of noise. Secondly, a self-attention gating mechanism was introduced in the skip connections of the Zero-DCE to enhance the network's attention to the cow area. Then, an improved kernel selection module was introduced in the feature fusion stage to adaptively adjust the size of the receptive field. Finally, a depthwise separable convolution was used to replace the standard convolution of Zero-DCE, and an Attentive Convolutional Transformer (ACT) module was used to replace the iterative approach in Zero-DCE, which further reduced the computational complexity of the network and speeded up the inference. Four different object-detection models, YOLOv5, CenterNet, EfficientDet, and YOLOv7-tiny, were selected to evaluate the performance of the improved network and were tested on the night-time dataset before and after enhancement. Experimental results demonstrate that the detection performance of all models is significantly improved when processing night-time image samples through the enhanced Zero-DCE model. In summary, the improved lightweight Zero-DCE low-light enhancement network proposed in this study shows excellent performance, which can ensure that various object-detection models can quickly and accurately identify targets in low-light environments at night and are suitable for real-time monitoring in actual production environments.



Citation: Yu, Z.; Guo, Y.; Zhang, L.; Ding, Y.; Zhang, G.; Zhang, D. Improved Lightweight Zero-Reference Deep Curve Estimation Low-Light Enhancement Algorithm for Night-Time Cow Detection. *Agriculture* **2024**, *14*, 1003. <https://doi.org/10.3390/agriculture14071003>

Academic Editor: Hyeon Tae Kim

Received: 28 May 2024

Revised: 18 June 2024

Accepted: 24 June 2024

Published: 26 June 2024

Keywords: image enhancement; low-light detection; precision livestock farming; Zero-DCE



Copyright: © 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Dairy farming plays a key role in the global economy, providing milk, meat, etc., to countries around the world [1,2], but it still faces many challenges. As the population grows and the demand for dairy products increases, livestock farming faces the question of how to improve productivity and animal health and welfare with limited resources. However, as the size of farms and livestock increases, farmers are unable to pay effective attention to individual animals, resulting in the inability to detect anomalies [3–6]. In this context, precision livestock farming has become an inevitable trend [7,8], so it is important to have technologies that can monitor animal information in real time.

Computer Vision Technology (CVT) has been widely used in pasture monitoring as a non-contact intelligent technology [9–12], and the rapid development of deep-learning technology has enabled CVT-based methods to obtain individual animal information and scene information quickly, accurately, and efficiently, and improve animal welfare [13,14], which is of great significance for the management decision-making of animal farming. Most studies were conducted under normal lighting conditions, ignoring the performance of the proposed methods to achieve target monitoring in low-light or night-time conditions. To achieve 24 h monitoring of farming, it is necessary to study target detection methods in low-light or night-time conditions. However, the image data generated under conditions such as night-time and low-light conditions often have problems with images being too dark or details being lost, which hinders the understanding of image content and the extraction of image features. Therefore, obtaining sufficient animal target features under these conditions is also a challenging task.

Image enhancement is an effective means of improving image quality. Traditional low-light enhancement algorithms include histogram equalization [15], Single Scale Retinex (SSR) [16], Naturalness Preserved Enhancement (NPE) [17], etc. However, the quality of image enhancement often depends on artificial prior knowledge and regularization, making it difficult to tune some complex situations. In addition, due to the complex optimization process of traditional low-light enhancement algorithms, the inference time is generally long, which is not suitable for real-time tasks.

With the development of deep learning, it has become a mainstream low-light image enhancement mode to establish the relationship between low-light input and enhanced output through heuristic design network structures and end-to-end learning. LightenNet [18] introduces a convolutional neural network into low-light image enhancement tasks and improves image quality through pre-processing and post-processing. However, it only enhances brightness and cannot process other image features. MBLEN [19] uses multiple subnets to enhance image features at different levels at the same time and fuses the output results of multiple branches to improve image quality from multiple aspects. However, this model requires a large amount of paired training data to perform optimally. EnlightenGAN [20] solves the problem of the difficulty in obtaining real paired training data to a certain extent by using Generative Adversarial Networks (GANs) for image enhancement, but it still needs to ensure that the training data have the same or similar data distribution, and the GAN-based method requires a large amount of computing resources. Guo et al. [21] proposed Zero-DCE (Zero-Reference Deep Curve Estimation), which eliminates the need for paired data and unpaired data through depth curve estimation and transforms the low-light enhancement problem into an image-specific curve estimation problem, which is faster in training. However, the Zero-DCE needs to keep the image size constant all the time during the training process, which makes it difficult for the network to handle the noise and brings challenges to the target detection task of night-time cows.

To meet the requirements of target detection for cows at night, it is not only necessary to process night-time images in real time but also to ensure the accuracy of object detection, so an improved lightweight Zero-DCE image enhancement model was proposed in this study. The main contributions of this work are as follows: it uses the upsampling structure in the original Zero-DCE to reduce noise impact, adds a self-attention gating mechanism for skip connections to make the network pay more attention to the cow area in the image, improves the kernel selection module to enhance multi-scale feature fusion, replaces the standard convolution operation of Zero-DCE with a depthwise separable convolution operation, and uses the ACT module to replace the iterative method of the original network, which reduces the computational cost of the network, making it lightweight. Finally, the night-time cow target detection was achieved by combining the current mainstream object-detection model.

The rest of the paper is organized as follows: Section 2 describes the research materials and methods, while Section 3 presents the results and analysis. Discussions are given in Section 4 with conclusions in Section 5.

2. Materials and Methods

2.1. Data Acquisition

The cow data were sourced from the Northwest Agriculture and Forestry University Animal Husbandry Experimental Base, Fufeng County, Baoji, Shaanxi Province, where cameras were used to capture cows in their natural states within actual feeding scenes in cattle pens [22]. Finally, nine video files, each 10 min long, were selected and converted into 1350 images at 100-frame intervals, and then the irrelevant parts at the top of the images were cropped in batches. Night-time light halos were processed to obtain night-time images. The collected image set includes cows at different times of the day and night, as shown in Figure 1.

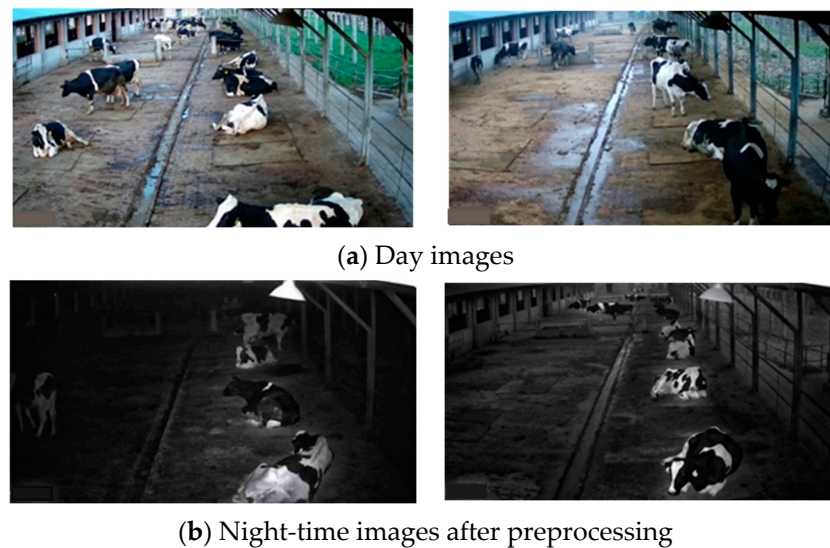


Figure 1. Examples of sample image dataset.

From the 1350 images obtained, 141 night-time and 559 day-time cow images, a total of 700 images, were selected to train the object-detection model. In addition, 154 images of night-time cows were selected and re-labeled as the test set of the images to be enhanced and the object-detection model.

2.2. Improved Lightweight Zero-DCE for Night-Time Cow Detection

2.2.1. Zero-DCE

Guo et al. [21] introduced the Zero-DCE network, which does not require reference images. Instead, it directly learns light enhancement curves through DCE-Net and utilizes these curves for image enhancement. The structure of DCE-Net is shown in Figure 2.

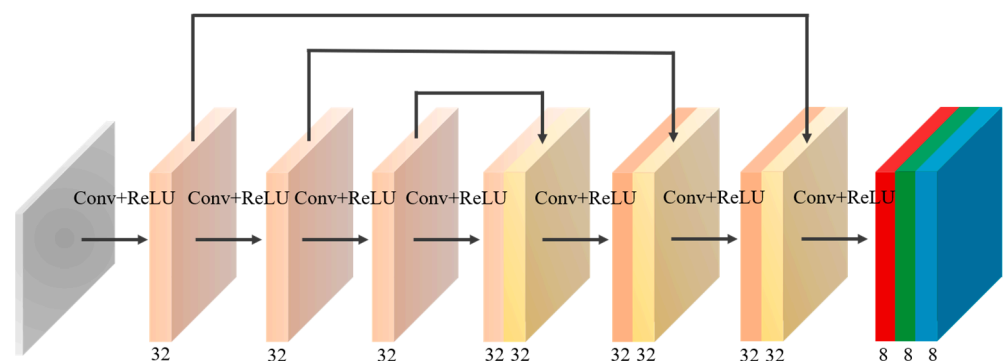


Figure 2. Structure of DCE-Net.

DCE-Net comprises a total of seven convolution operations. During the 1st to the 4th convolution operation, the input image channels are expanded and image features are extracted. Before the 5th to 7th convolution operations, skip connections are performed to expand the number of channels to further enhance feature extraction. Finally, channel compression is performed through convolutional and activation functions, yielding an output with feature maps containing eight dimensions for each of the RGB channels.

Specifically, the steps of the Zero-DCE algorithm are as follows: First, it converts the input low-light image into a grayscale image and standardizes it. Then, the algorithm utilizes a deep curve estimation network (DCE-Net) to generate a light-enhancement curve, A_n (obtained from the feature graph segmentation of the final output of DCE-Net), ensuring that the generated curve matches the features of the original image by minimizing the contrast loss function. Finally, the original image is pixel-wise adjusted using the generated light-enhancement curves through multiple iterations, ultimately producing the enhanced image. The iterative process of pixel-wise adjustment can be expressed by the following formula,

$$LE_n(x) = LE_{n-1}(x) + A_n(x)LE_{n-1}(x)(1 - LE_{n-1}(x)), \quad (1)$$

where x is the pixel coordinate, $LE_n(x)$ is the enhancement result of the input image $LE_{n-1}(x)$, A_n is the curve parameter chart, which is consistent with the dimension of the input image, and n is the number of iterations, where n is 8. The overall framework of this network is shown in Figure 3.

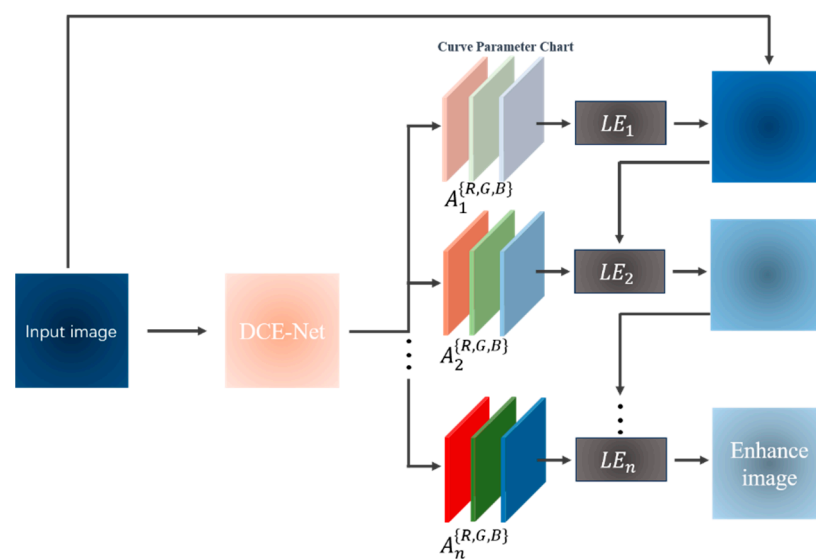


Figure 3. Framework of DCE-Net.

2.2.2. Self-Attention Gate Mechanism

Shallow networks often contain many redundant features, and skip connections require concatenating feature maps from shallow and deep networks, introducing noise in the feature extraction process. The Attention Gate (AG) can implicitly learn to suppress irrelevant regions in the input image while emphasizing salient features useful for specific tasks. To enable the feature extraction network to capture more critical task-specific information, this study extends the AG and introduces a self-attention gate mechanism in the skip connection. The overall structure of the self-attention gate mechanism is shown in Figure 4.

Firstly, the $32 \times H \times W$ feature map of the deep network undergoes channel compression through a 1×1 convolution operation, followed by an ReLU activation operation to activate the grid signal of spatial information in the image. This activation allows the skip connection to fuse features of different scales, enriching semantic information. Next,

the obtained $16 \times H \times W$ feature map and the $16 \times \frac{H}{4} \times \frac{W}{4}$ feature map from the shallow network undergo channel expansion and size transformation through convolution kernels of size 1×1 and 6×6 , respectively. These transformed feature maps are then element-wise added to create a new $32 \times \frac{H}{4} \times \frac{W}{4}$ feature map. The new feature map is subjected to ReLU activation, a 1×1 convolutional kernel, and feature values are kept in the range of 0 to 1. Subsequently, the final attention map is obtained through upsampling. Finally, the feature map from the shallow network is multiplied element-wise with the $16 \times H \times W$ attention map, thus combining the contextual information extraction ability of the coarse-grained feature map with the texture information extraction ability of the fine-grained feature map to generate the final feature map.

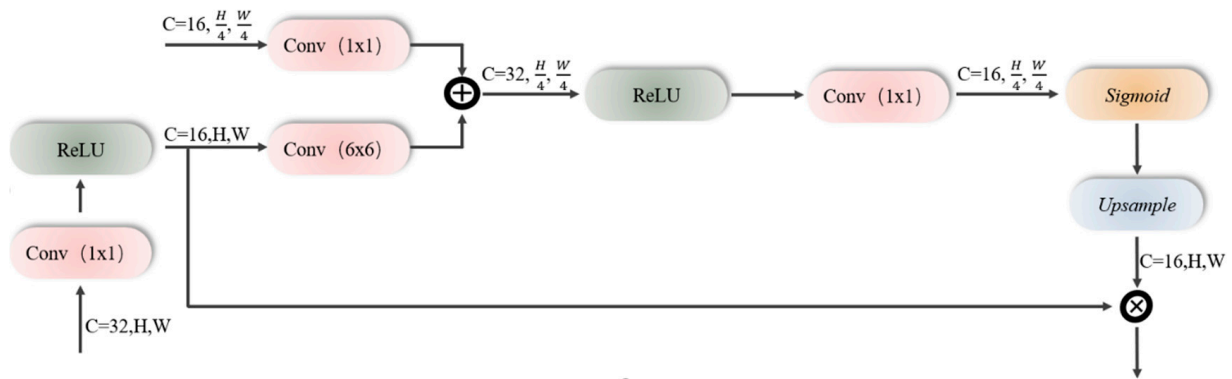


Figure 4. Structure of self-attention gate mechanism.

2.2.3. Kernel Selection Module

When the image is heavily affected by strong noise, local information around the noise is lost. Therefore, it is crucial to explore contextual information around the noise. As a lightweight operation, Selective Kernel (SK) convolution can adaptively adjust its receptive field size, fully utilizing spatial features for multi-scale feature fusion, enhancing the efficiency and effectiveness of object recognition. To address the continuously changing noise distribution in night-time cow images, this study extends the SK convolution dual-branch and introduces it into the feature fusion stage. The overall structure of the improved kernel selection module is shown in Figure 5.

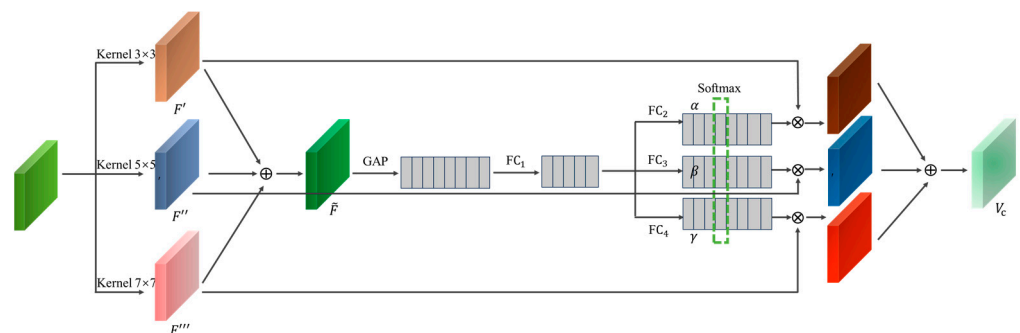


Figure 5. Structure of kernel selection module.

Specifically, for the input feature map, F , convolution operations with kernel sizes of 3, 5, and 7 are applied, followed by batch normalization and ReLU activation, resulting in corresponding feature maps F' , F'' , and F''' . To improve efficiency, the conventional convolution with 5×5 and 7×7 kernels is replaced with a 3×3 convolution and dilated convolution with the appropriate dilation size. The three feature maps are then summed to obtain $\tilde{F}$. Through Global Average Pooling (GAP) and Fully Connected (FC) layering, compression and expansion are performed to obtain three weight coefficients, α , β , and γ .

Subsequently, using the Softmax activation function, the corresponding weights, k_c , for each channel are obtained:

$$k_c = \frac{e^{k_c}}{e^{\alpha_c} + e^{\beta_c} + e^{\gamma_c}}, k = \alpha, \beta, \gamma, \quad (2)$$

where, α , β , and γ represent the soft attention vectors for the feature maps F' , F'' , and F''' , with α_c denoting the c -th element of α , and similarly for β_c and γ_c . Finally, the feature maps processed with different-sized convolution kernels are multiplied by their respective soft attention vectors to obtain the final output feature map, V_c :

$$V_c = \alpha_c F' + \beta_c F'' + \gamma_c F''' \quad (3)$$

2.2.4. Depthwise Separable Convolution

On devices with limited computing resources, it is challenging for conventional convolution operations to achieve real-time performance due to their high computational demands. Depthwise (DW) convolution combined with Pointwise (PW) convolution is employed in depthwise separable convolution. This structure, like regular convolution, is used for feature extraction. However, in comparison to conventional convolution, depthwise separable convolution has a lower number of parameters and computational costs, resulting in faster real-time execution. Such structures are commonly employed in various lightweight networks, such as MobileNet [23], ShuffleNet [24], etc. The structure of depthwise separable convolution is shown in Figure 6.

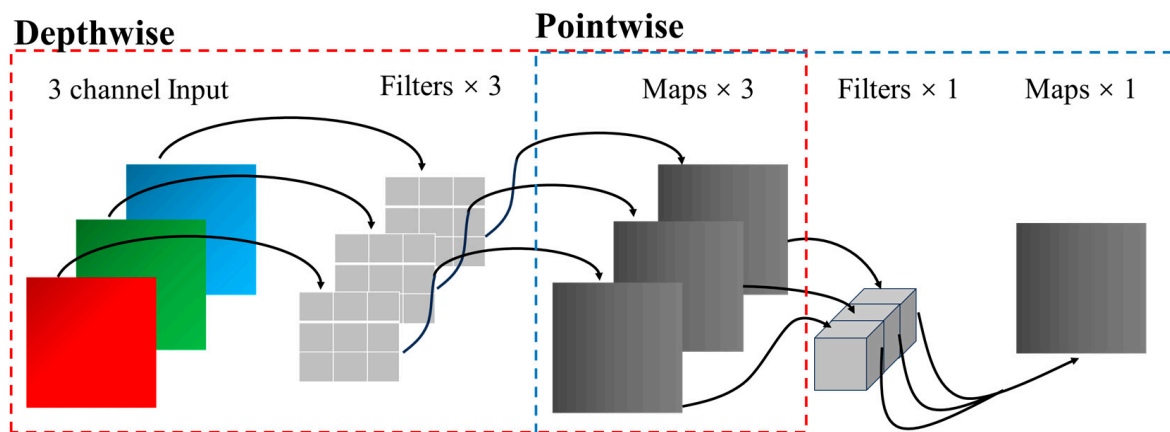


Figure 6. Structure of depthwise separable convolution.

(1) Depthwise

Depthwise convolution adopts a separate approach, where each convolutional kernel is responsible for processing one input channel. In other words, each channel is convolved only with one kernel. In contrast, in regular convolution, each kernel simultaneously operates on all channels of the input image.

Taking a three-channel color input image as an example, depthwise convolution first performs convolution separately with three kernels, one for each channel. This means that channels and convolutional kernels are one-to-one corresponded. Therefore, after this operation, a three-channel image generates three feature maps. However, the number of feature maps after depthwise convolution is the same as the number of channels in the input layer, and it cannot expand the number of feature maps. Additionally, this operation independently convolves each channel of the input layer, without fully utilizing the feature information of different channels at the same spatial position. Hence, pointwise convolution is needed to combine these feature maps and generate new feature maps.

(2) Pointwise

The operation of pointwise convolution is similar to regular convolution, with a kernel size of $1 \times 1 \times M$, where M is the number of channels in the previous layer. In this convolution operation, the feature maps are linearly combined along the channel direction to generate new feature maps, and the number of output feature maps is equal to the number of kernels used. In summary, depthwise separable convolution is accomplished through two consecutive convolution operations.

The original Zero-DCE feature extraction network uses regular convolution operations. To further meet the real-time requirements, this study replaces regular convolutions with depthwise separable convolutions to significantly reduce the network's parameter count and accelerate the inference speed.

2.2.5. ACT Module

The original Zero-DCE network suffers from slow processing speed, primarily due to the extensive computation involved in the eight iterations of curve adjustment. To address this, this study introduces the ACT module, which replaces the original eight iterations with a convolution, normalization, and activation process. The structure of the ACT module is shown in Figure 7.

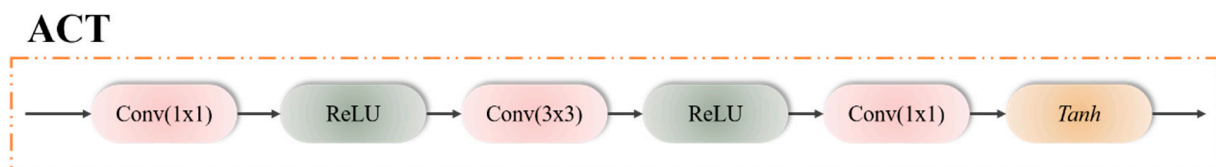


Figure 7. Structure of ACT module.

The activation module consists of three convolutional layers interleaved with three activation layers. The first two activation layers use the ReLU function, and the third activation layer uses the tanh function, enhancing low-light features. This results in a set of enhancement coefficients used for the final enhancement curve. This approach alleviates the issue of redundant iterations, significantly reducing computational complexity.

2.2.6. Improved Lightweight Zero-DCE

In this study, we proposed a night-time cow image enhancement detection framework based on an improved lightweight Zero-DCE network. In addition to the above modifications, this approach replaces the feature extraction network of the original Zero-DCE network with an up-sampling and down-sampling structure to suppress noise and reduce external environmental interference. The overall process consists of two steps: (1) inputting night-time, low-light cow images into the improved low-light enhancement network for enhancement; (2) detecting the obtained enhanced images using an object-detection model. The overall structure of the improved lightweight Zero-DCE network is shown in Figure 8.

This study uses night-time cow images as input for the improved Zero-DCE. The specific process is as follows:

An image with dimensions $H \times W \times 3$ is input into the network. First, channel expansion and downsampling are performed through three convolution layers and three activation function operations, resulting in a feature map of dimensions $H/16 \times W/16 \times 16$. Subsequently, two upsampling operations are applied to restore the feature map to dimensions $H \times W \times 48$. Simultaneously, skip connections are introduced in the upsampling operations, concatenating the feature maps of the shallow network and the deep network along the channel direction to retain spatial information from the shallow network. The self-attention gate mechanism is employed to suppress redundant features in the shallow network. The restored feature map undergoes channel compression through convolution and activation function operations. It is then input into the kernel selection module to obtain an output feature map of dimensions $H \times W \times 32$. The output feature map is

processed by the ACT module to obtain the final curve parameter map. Finally, the image is enhanced using the curve parameter map. The enhancement formula is as follows:

$$LE(I(x); A_n) = I(x) + A_n I(x)(1 - I(x)), \quad (4)$$

where x represents the pixel coordinates of the image, $I(x)$ is the input image, $LE(I(x); A_n)$ is the enhanced image of the input, and A_n represents the curve parameter map.

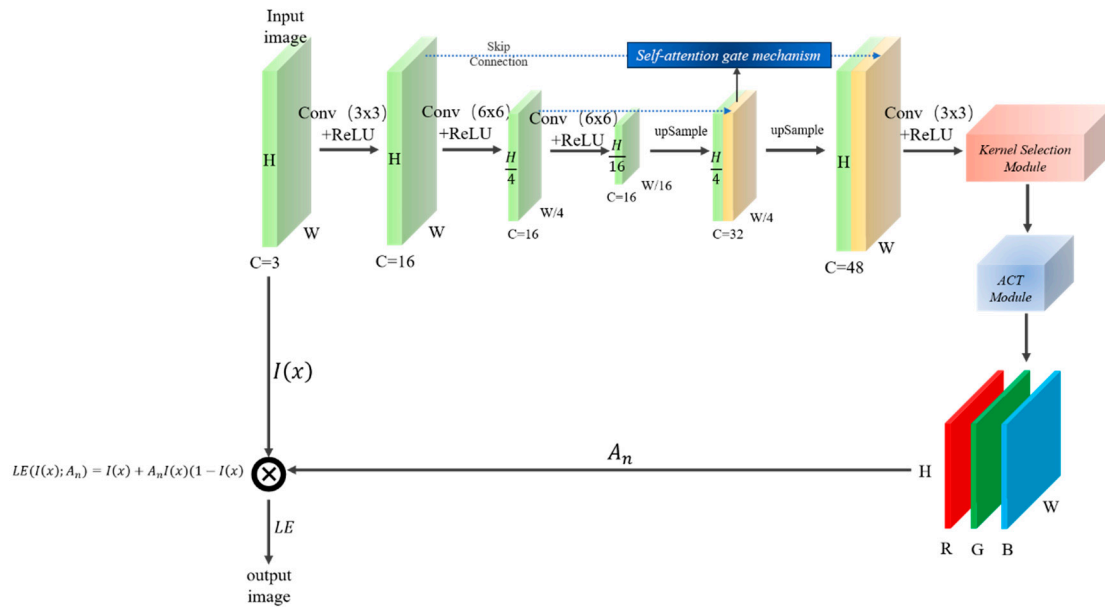


Figure 8. Structure of improved lightweight Zero-DCE.

The improved network still uses the loss function of Zero-DCE, mainly including spatial consistency loss (L_{spa}), exposure control loss (L_{exp}), color constancy loss (L_{col}), and illumination smoothness loss (L_{tv_A}). The overall loss expression is as follows:

$$L_{total} = \lambda_1 L_{spa} + \lambda_2 L_{exp} + \lambda_3 L_{col} + \lambda_4 L_{tv_A}. \quad (5)$$

Spatial consistency loss is used to prevent significant changes in the value of a pixel and its neighboring pixels before and after image enhancement. The specific formula is as follows:

$$L_{spa} = \frac{1}{K} \sum_{i=1}^K \sum_{j \in \Omega(i)} (|Y_i - Y_j| - |I_i - I_j|)^2, \quad (6)$$

where K represents the number of pixels, $\Omega(i)$ represents the four neighborhoods (up, down, left, right) of the i -th pixel, and Y and I denote the local region average values for the enhanced and input images, respectively.

Exposure control loss represents the distance between the brightness of each pixel and some intermediate value, aiming to prevent certain areas from being too dark or too bright. The specific formula is as follows:

$$L_{exp} = \frac{1}{M} \sum_{k=1}^M |Y_k - E|, \quad (7)$$

where E represents the ideal intermediate brightness value, set to 0.6 in this paper, Y denotes the local region brightness average of the enhanced image, and M represents the number of non-overlapping 16×16 regions.

Color constancy loss is based on the conclusion that the values of a color channel in an image should not significantly exceed those of other channels [25]. A correction relationship

is established among the three RGB channels to ensure that the average values of the RGB channels in the enhanced image are as close as possible. The specific formula is as follows:

$$L_{col} = \sum_{(p,q) \in \varepsilon} (J_p - J_q)^2, \varepsilon = \{(R, G), (R, B), (G, B)\}, \quad (8)$$

where, (p, q) traverses all pairwise combinations of the three-color channels and J_p and J_q represent the average brightness values for color channels p and q , respectively.

Illumination smoothness loss is designed to maintain a monotonic relationship between adjacent pixels, ensuring that the brightness changes between neighboring pixels are not too pronounced. The specific formula is as follows:

$$L_{tv_A} = \frac{1}{M} \sum_{n=1}^N \sum_{c \in \xi} (|\nabla_x A_n^c| + |\nabla_y A_n^c|)^2, \xi = \{R, G, B\}. \quad (9)$$

In the above equation, N represents the number of iterations, ∇_x and ∇_y are the horizontal and vertical gradient operators, and A_n^c represents the curve parameter map for each channel.

2.3. Evaluation Indicators

The aim of this study is to enhance the accuracy and robustness of night-time, low-light detection in cattle. Therefore, precision, recall, mean average precision (mAP), and frames per second (FPS) are employed as evaluation metrics for the network, as shown in the following equations:

$$precision = \frac{TP}{TP + FP} \times 100\%, \quad (10)$$

$$recall = \frac{TP}{TP + FN} \times 100\%, \quad (11)$$

$$AP = \int_0^1 P(r) dr, \quad (12)$$

$$mAP = \frac{1}{n} \sum_{i=1}^n AP(i) \quad (13)$$

where TP , FP , and FN represent the true positive, false positive, and false negative sample counts, respectively. mAP is the average precision across all classes. Additionally, n denotes the number of classification types (in this study, $n = 1$). FPS refers to the number of enhanced images inferred within one second.

3. Results and Analysis

All experiments in this study were conducted in a Microsoft Windows 11 64-bit system environment, using the PyTorch-1.12.1 version, an open-source deep-learning framework that supports GPU acceleration and dynamic neural networks. The experiments were conducted with CUDA 11.6 for training. The computer (Lenovo Legion R9000P 2021H, China) was used for the experiments, and its own hardware information is as follows: AMD Ryzen 7 5800H with Radeon Graphics@3.20 GHz processor, 16 GB of RAM, and an RTX 3070 graphics card. Additionally, to expedite model convergence, pre-trained models corresponding to each detection algorithm were utilized as the initial weights for the networks.

3.1. Comparison of the Results of Different Models

The comparative experiments were conducted to better showcase the performance of the improved enhancement model. In this study, the detection results on the original test set and the test set enhanced by the improved enhancement model were compared across four different object-detection models: YOLOv5, CenterNet, EfficientDet, and YOLOv7-tiny. The models were trained and validated using the same dataset of cow images. Table 1

presents the comparative results of precision, recall, mAP, and the inference time for the improved enhancement model across the various models.

Table 1. Target detection results before and after night image enhancement.

Models	Unenhanced			Improved Zero-DCE Enhanced			
	P	R	mAP0.5	P	R	mAP0.5	FPS (Frame/s)
YOLOv5	92.6	70.8	83.1	94.7	80.1	88.4	346.5
CenterNet	94.9	67.95	75.09	92.43	71.81	79.23	
EfficientNet	87.17	74.24	83.45	91.51	78.7	85.95	
YOLOv7-tiny	91.6	79.2	78.7	92.9	79.8	80.9	

As can be seen from Table 1, the recall rate of YOLOv5 on the enhanced test set is as high as 80.1%, which is 9.3% higher than that of the original test set, and other indicators are also improved, with an accuracy rate of 94.7% and mAP@0.5 being 83.1%, which are 2.1% and 5.3%, respectively. Although CenterNet's accuracy on the enhanced test set was slightly lower than that of the original test set, the recall and mAP@0.5 rates increased by 3.86% and 4.14%, respectively. Compared with the original test set, the accuracy, recall and mAP@0.5 of EfficientNet on the enhanced test set are increased by 4.34%, 4.46%, and 2.5%, respectively. YOLOv7-tiny also improved the accuracy, recall, and mAP@0.5 by 1.3%, 0.6%, and 2.2%, respectively. In summary, among the four object-detection models, all the indicators of all models on the enhanced test set are higher than those of the original test set, except that CenterNet's accuracy on the enhanced test set is slightly lower than that of the original test set. Therefore, the improved Zero-DCE network can enhance the night-time cow images and improve the detection performance of the detection model at night. In addition, its FPS is up to 346.5 f/s, which fully meets the speed requirements of real-time processing.

3.2. Ablation Experiment

In order to better validate the effectiveness of each improvement module in optimizing the original Zero-DCE network, we selected the YOLOv5 object-detection model with the best detection results and the maximum improvement on the enhanced test set as the baseline for ablation experiments. A total of five network configurations were tested using the same night-time cow test set, and the experimental results are shown in Table 2.

Table 2. Ablation experiments.

Model	DCE+ Improvement Modules				Detection Results after Image Enhancement		
	Up-Down Sampling Structure	Self-Attention Gate Mechanism	Kernel Selection Module	ACT Module	P	R	mAP0.5
YOLOv5	×	×	×	×	90.8	74.2	84.7
	✓	×	×	×	93.5	72.8	84.7
	✓	✓	×	×	94.0	73.4	85.1
	✓	✓	✓	×	93.0	75.1	85.9
	✓	✓	✓	✓	94.7	80.1	88.4

Note: × indicates that the module is not used, and ✓ indicates that the module is used.

From the data in Table 2, it can be seen that after replacing the feature extraction network in Zero-DCE with the feature extraction network with the up-and-down sampling structure, the accuracy rate increased by 2.7 percentage points, and the mAP@0.5 did not change, but the recall rate decreased slightly. After the introduction of the attention mechanism, the accuracy rate was further improved, and the recall rate rebounded, with

mAP@0.5 increasing by 0.4 percentage points. After adding the core selection module, the accuracy, recall, and mAP@0.5 indicators are better than those of the original Zero-DCE network, which are improved by 2.2%, 0.9%, and 1.2%, respectively. Finally, after eight iterations of the ACT module instead of the seventh shallow feature layer of Zero-DCE, the three indexes were further improved, and the accuracy rate was 94.7%. The recall rate was 80.1% and mAP@0.5 was 88.4%, which were 3.9%, 5.9%, and 3.7% higher than the original Zero-DCE, respectively. In summary, on the whole, with the introduction of each improvement module, the three evaluation indicators have been improved; in particular, mAP@0.5 shows a steady and increasing trend. In order to illustrate the superiority of the improved Zero-DCE network over the original Zero-DCE network, the YOLOv5 model was further used to predict the images of the same night-time dairy cow test set after the enhancement of the two networks, and the comparison and prediction images are shown in Figure 9.

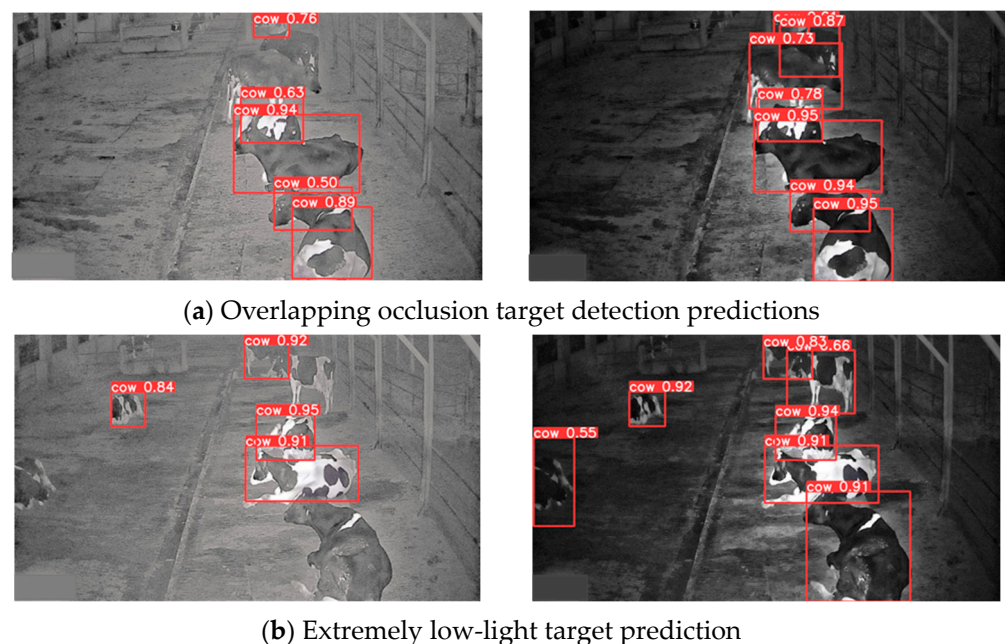


Figure 9. Comparison of partial prediction results between Zero-DCE and improved Zero-DCE.

On the left is the prediction image enhanced by the original Zero-DCE network, and on the right is the prediction image enhanced by the improved Zero-DCE network. In Figure 9a is the detection of the occluded target; it can be seen that the occluded target is missed in the left figure, while the occluded target is detected in the right figure, but there is still a false detection area; Figure 9b, in order to detect very low-illuminated targets, there are many missed detections in the left figure, and the detection is normal in the right figure. In addition, overall, the prediction frame on the right fits more accurately than the detection frame on the left. In summary, the improved Zero-DCE network has a superior enhancement effect.

3.3. Detection Results before and after Day-Time Data Enhancement

In order to verify that the augmentation model proposed in this paper is combined with the YOLOv5 object-detection model to achieve object detection during the day and at night, it is further verified on the day-time dataset. The test results are shown in Figure 10. It can be seen that the color tone of the enhanced image has changed, but object detection can still be achieved. Therefore, the improved Zero-DCE model proposed in this paper is combined with the YOLOv5 object-detection model to achieve day-time and night-time target detection once the network training is completed.

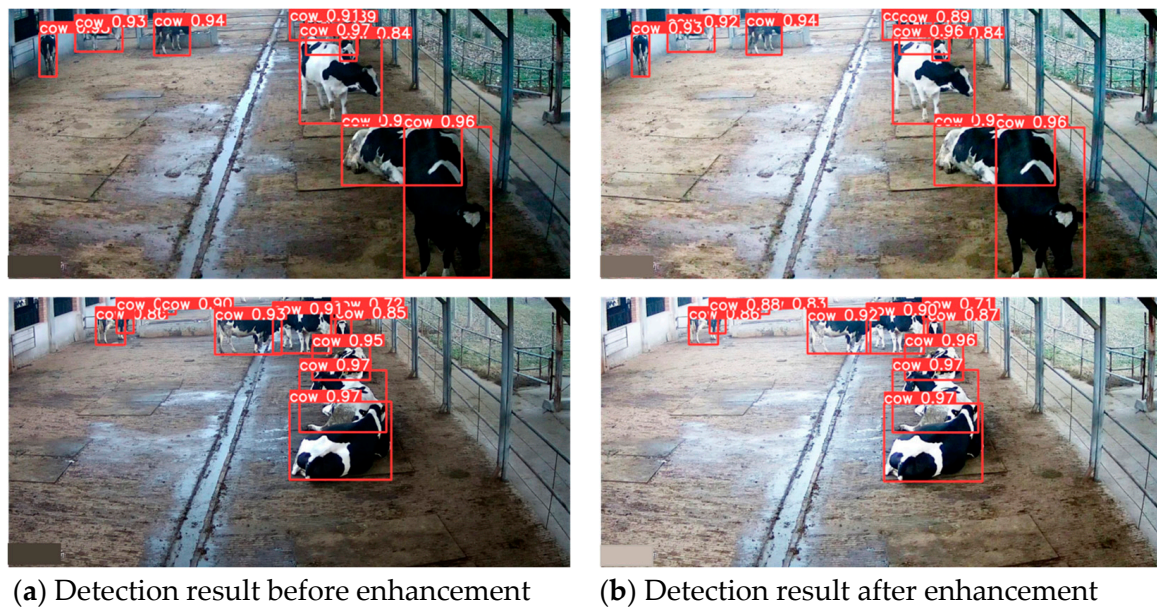


Figure 10. Cow target detection results in day-time images before and after enhancement.

4. Discussion

4.1. The Influence of Datasets on the Experimental Results

The datasets used in this study were all obtained from surveillance videos, and most of the night-time images are of poor quality, such as there being more image noise or low resolution, due to insufficient light and more environmental interference at night. The performance of the original Zero-DCE is greatly affected by the quality of the input images and the number of low-light images in the training dataset. Therefore, in practical applications, the training dataset can be further expanded by adding a large number of high-quality, low-light images and performing multiple exposure processing to improve the performance and applicability of the improved Zero-DCE network.

4.2. Performance Analysis of Improved Lightweight Zero-DCE

The improved Zero-DCE network proposed in this study achieves a precision rate of 94.7% on YOLOv5, surpassing the original Zero-DCE network. The up-sampling structure and kernel selection module help suppress noise and reduce external environmental interference, and the self-attention gating mechanism enhances the weight of the task targets throughout the feature map, leading to better enhancement and improved subsequent detection results. The ACT module and depthwise separable convolution significantly reduce the network's computational load, increasing the inference speed to meet real-time requirements. From Tables 1 and 2, as well as the accompanying figures, it is evident that the improved Zero-DCE network not only enhances the original Zero-DCE's performance but also boosts overall night-time detection capabilities across different object-detection models like YOLOv5, CenterNet, EfficientNet, and YOLOv7-tiny. However, even after enhancement with the improved Zero-DCE network, there are occasional false negatives and false positives when the edges of the image are too dark, noise levels are high, and cows are visibly overlapping or occluded.

4.3. Possible Research Directions in the Future

Although this study proposed an effective enhancement framework for night-time cow detection to improve the accuracy of detection, researchers still need to address a series of limitations and challenges. Therefore, in future research, on the one hand, we can start from multiple scenes and no longer focus on a single application scene. On the other hand, we can start with the dataset and look at how to obtain a higher quality dataset, or investigate what pre-processing measures should be taken to improve the image quality.

5. Conclusions

In this study, the improved lightweight Zero-DCE low-light enhancement network is proposed to enhance the detection of dairy cow night-time images, which successfully integrates the up-down sampling structure, self-attention gating mechanism, kernel selection module, ACT module, and deep separable convolution into Zero-DCE, which improves its detection performance on the object-detection model, and especially improves the problem that the network has difficulty dealing with noise and has large computational cost. The accuracy rate, recall rate, and mAP@0.5 of the improved lightweight Zero-DCE network on YOLOv5 reached 94.7%, 80.1%, and 88.4%, respectively, which is better than the original Zero-DCE network and is suitable for a variety of different object-detection models. The overall network size is small and easy to deploy to embedded devices, and the FPS reaches 346.5 f/s, which is far more than enough to meet the needs of real-time enhanced processing. In conclusion, the night-time image enhancement detection method of dairy cows based on deep-learning technology proposed in this paper is conducive to the long-term, real-time animal detection of precision animal farming.

Author Contributions: Z.Y. and Y.G., methodology, validation, formal analysis, and writing—original draft; L.Z., conceptualization, methodology (lead), formal analysis, and writing—review and editing; Y.D., writing—review and editing; G.Z., writing—review and editing; D.Z., resources, funding acquisition, and writing—review and editing. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported by the Natural Science Foundation of Anhui Higher Education Institutions of China (No. 2023AH050082), the Science and Technology Project of Inner Mongolia Autonomous Region (No. 2022YFSJ0039), the Open Fund of the Key Laboratory of Agricultural Internet of Things of the Ministry of Agriculture (No. 2023AIOT-04), the Key Scientific Research Projects of Anhui Provincial Department of Education (No. 2022AH051373), and Agricultural Ecological Big Data Analysis and Application Technology National Local Joint Engineering Center of Anhui University (No. AE202201xp).

Data Availability Statement: The original data can be provided by the corresponding author upon reasonable request.

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

References

1. Khan, N. Critical Review of Dairy Cow Industry in the World. *Food Chemistry eJournal*. 2020. Available online: <https://api.semanticscholar.org/CorpusID:219396510> (accessed on 23 June 2024).
2. Wang, L.; Wu, T.; Zhang, Y.; Yang, K.; He, Y.; Deng, K.; Liang, C.; Gu, Y. Comparative studies on the nutritional and physicochemical properties of yoghurts from cows', goats', and camels' milk powder. *Int. Dairy J.* **2023**, *138*, 105542. [\[CrossRef\]](#)
3. Guo, Y.; Hong, W.; Wu, J.; Huang, X.; Qiao, Y.; Kong, H. Vision-Based Cow Tracking and Feeding Monitoring for Autonomous Livestock Farming: The YOLOv5s-CA+DeepSORT-Vision Transformer. *IEEE Robot. Autom. Mag.* **2023**, *30*, 68–76. [\[CrossRef\]](#)
4. Han, S.; Fuentes, A.; Yoon, S.; Jeong, Y.; Kim, H.; Park, D.S. Deep learning-based multi-cattle tracking in crowded livestock farming using video. *Comput. Electron. Agric.* **2023**, *212*, 108044. [\[CrossRef\]](#)
5. Zheng, Z.; Li, J.; Qin, L. YOLO-BYTE: An efficient multi-object tracking algorithm for automatic monitoring of dairy cows. *Comput. Electron. Agric.* **2023**, *209*, 107857. [\[CrossRef\]](#)
6. Qiao, Y.; Guo, Y.; He, D. Cattle body detection based on YOLOv5-ASFF for precision livestock farming. *Comput. Electron. Agric.* **2023**, *204*, 107579. [\[CrossRef\]](#)
7. Jin, H.; Meng, G.; Pan, Y.; Zhang, X.; Wang, C. An improved intelligent control system for temperature and humidity in a pig house. *Agriculture* **2022**, *12*, 1987. [\[CrossRef\]](#)
8. Mazzetto, A.M.; Falconer, S.; Ledgard, S. Mapping the carbon footprint of milk production from cattle: A systematic review. *J. Dairy Sci.* **2022**, *105*, 9713–9725. [\[CrossRef\]](#) [\[PubMed\]](#)
9. Wang, X.; Chen, B.; Yang, R.; Liu, K.; Cuan, K.; Cao, M. A Non-Contact and Fast Estimating Method for Respiration Rate of Cows Using Machine Vision. *Agriculture* **2023**, *14*, 40. [\[CrossRef\]](#)
10. Xu, B.; Cui, X.; Ji, W.; Yuan, H.; Wang, J. Apple grading method design and implementation for automatic grader based on improved YOLOv5. *Agriculture* **2023**, *13*, 124. [\[CrossRef\]](#)
11. Ji, W.; Pan, Y.; Xu, B.; Wang, J. A real-time apple targets detection method for picking robot based on ShufflenetV2-YOLOX. *Agriculture* **2022**, *12*, 856. [\[CrossRef\]](#)

12. Hu, T.; Wang, W.; Gu, J.; Xia, Z.; Zhang, J.; Wang, B. Research on Apple Object Detection and Localization Method Based on Improved YOLOX and RGB-D Images. *Agronomy* **2023**, *13*, 1816. [\[CrossRef\]](#)
13. Wang, Z.; Wang, S.; Wang, C.; Zhang, Y.; Zong, Z.; Wang, H.; Su, L.; Du, Y. A non-contact cow estrus monitoring method based on the thermal infrared images of cows. *Agriculture* **2023**, *13*, 385. [\[CrossRef\]](#)
14. Chen, C.; Zhu, W.; Steibel, J.; Siegford, J.; Han, J.; Norton, T. Classification of drinking and drinker-playing in pigs by a video-based deep learning method. *Biosyst. Eng.* **2020**, *196*, 1–14. [\[CrossRef\]](#)
15. He, K.M.; Sun, J.; Tang, X.O. Single image haze removal using dark channel prior. *IEEE Trans. Pattern Anal.* **2011**, *33*, 2341–2353.
16. Jobson, D.J.; Rahman, Z.; Woodell, G.A. Properties and performance of a center/surround Retinex. *IEEE Trans. Image Process.* **1997**, *6*, 451–462. [\[CrossRef\]](#) [\[PubMed\]](#)
17. Wang, S.H.; Zheng, J.; Hu, H.M.; Li, B. Naturalness preserved enhancement algorithm for non-uniform illumination images. *IEEE Trans. Image Process.* **2013**, *22*, 3538–3548. [\[CrossRef\]](#) [\[PubMed\]](#)
18. Li, C.; Guo, J.; Porikli, F.; Pang, Y. LightenNet: A convolutional neural network for weakly illuminated image enhancement. *Pattern Recogn. Lett.* **2018**, *104*, 15–22. [\[CrossRef\]](#)
19. Lv, F.; Lu, F.; Wu, J.; Lim, C. MBLEN: Low-Light Image/Video Enhancement Using CNNs. *BMVC* **2018**, *220*, 4.
20. 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. [\[CrossRef\]](#)
21. Guo, C.; Li, C.; Guo, J.; Loy, C.C.; Hou, J.; Kwong, S.; Cong, R. Zero-reference deep curve estimation for low-light image enhancement // Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Seattle, WA, USA, 14–19 June 2020; pp. 1780–1789.
22. Guo, Y.; Zhang, Z.; He, D.; Niu, J.; Tan, Y. Detection of cow mounting behavior using region geometry and optical flow characteristics. *Comput. Electron. Agric.* **2019**, *163*, 104828. [\[CrossRef\]](#)
23. Howard, A.G.; Zhu, M.; Chen, B.; Kalenichenko, D.; Wang, W.; Weyand, T.; Andreetto, M.; Adam, H. Mobilenets: Efficient convolutional neural networks for mobile vision applications. *arXiv* **2017**, arXiv:1704.04861.
24. Zhang, X.; Zhou, X.; Lin, M.; Sun, J. Shufflenet: An extremely efficient convolutional neural network for mobile devices. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–22 June 2018; pp. 6848–6856.
25. Buchsbaum, G. A spatial processor model for object colour perception. *J. Frankl. Inst.* **1980**, *310*, 1–26. [\[CrossRef\]](#)

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