Impact of Adverse Lighting Conditions on Deep Learning Object Detection

Authors: Liu Jingshuo

Date: 2024-01-09T00:00:00+00:00

Abstract

Object detection under adverse lighting conditions is an important image processing task. Current research primarily reduces image noise through image enhancement, while simultaneously improving network structures and datasets to adapt to object detection under adverse lighting conditions. However, few studies have investigated the specific effects of adverse lighting conditions on object detection. Therefore, in this paper, we generate datasets simulating adverse lighting conditions through algorithms, perform object detection under different noise conditions, and statistically analyze the detection results to study these effects. Since the experiments are conducted on simulated data, to ensure the accuracy of the results, we validate the conclusions using real-world images captured under adverse lighting conditions.

Full Text

Abstract

Object detection under adverse lighting conditions represents an important image processing task. Current research primarily addresses this challenge through image enhancement techniques to reduce noise, alongside improvements to network architectures and datasets. However, few studies have systematically investigated the specific effects of adverse lighting conditions on object detection performance. In this paper, we generate simulated adverse lighting datasets through algorithmic approaches and conduct object detection experiments across varying noise conditions. By statistically analyzing the detection results, we quantify and investigate these impacts.

Keywords: Low illumination, YOLOv5, object detection

1. Research Background

1.1 Research Significance

In recent years, autonomous driving has emerged as a prominent research area attracting considerable attention from researchers. In this domain, object detection plays a crucial role in ensuring safe driving. Two-dimensional object detection receives camera data to predict surrounding objects and graphical information such as traffic signs and crosswalks. With advances in deep learning, the accuracy and speed of 2D object detection have continuously improved. Current object detection algorithms mainly fall into two categories: One-Stage and Two-Stage, with representative algorithms including the YOLO series, RCNN series, and Retina-Net. Under normal lighting conditions, 2D object detection has already achieved high accuracy.

However, real-world environmental complexity cannot guarantee consistent image input quality. Under adverse lighting conditions, feature information in input images becomes blurred, which significantly impacts object detection performance. In the autonomous driving domain, achieving all-weather operation of intelligent systems makes object detection under adverse lighting conditions an essential problem to solve.

Numerous studies have addressed object detection under adverse lighting conditions. Many researchers have attempted to reduce noise in such images through image enhancement and preprocessing methods, including adjusting image contrast, enhancing brightness, and removing noise to make targets more detectable. Additionally, deep learning models have achieved remarkable progress in solving this problem, as they can learn features from large datasets and exhibit certain robustness to lighting variations. Meanwhile, some studies have fused visible light images with data from other sensors, such as infrared and thermal imaging, to improve detection performance across different lighting conditions. Multi-modal fusion helps provide more comprehensive and reliable information.

Nevertheless, few researchers have analyzed the magnitude of impact that adverse lighting conditions exert on object detection. Therefore, quantifying these effects is necessary. This research enables researchers to more intuitively understand how image noise affects object detection, thereby helping them develop more reasonable improvement strategies tailored to different scenarios.

1.2 Research Status

Regarding object detection methods, current approaches fall into two categories: One-Stage and Two-Stage, with many excellent models available. One-Stage methods directly extract features for object detection, offering fast speed and avoiding background false positives, but suffer from lower detection rates and localization accuracy. Main algorithms include the YOLO series, SSD series, Retina-Net, and DetectNet. Two-Stage methods first generate candidate regions before performing classification and localization, featuring longer training times

and slower speed but higher detection precision. Main algorithms include the RCNN series, SPPNet, and R-FCN.

For adverse lighting object detection, the primary solution involves using low-illumination image enhancement algorithms. Common enhancement methods include histogram equalization, homomorphic filtering, and Retinex theory. With deep learning development, deep learning-based enhancement methods have also matured. Improving detection networks and using adverse lighting datasets can further enhance detection capabilities under such conditions.

2. Related Theory and Technology

2.1.1 RCNN Algorithm

RCNN is a two-stage object detection algorithm based on convolutional neural networks and represents the first successful CNN application to object detection. Its core idea involves classifying candidate regions. First, a selection algorithm extracts candidate regions, each of which undergoes feature extraction through a CNN. Subsequently, a support vector machine classifies each region. For candidate boxes classified as positive samples, a regressor fine-tunes them to obtain more accurate detection boxes. Compared with traditional algorithms, RCNN achieves higher accuracy but suffers from slow computational speed and substantial resource consumption. Moreover, traditional CNNs require fixed-size inputs, causing information loss during image scaling. To address this, researchers proposed Spatial Pyramid Pooling Networks (SPPnet), eliminating the need for fixed input sizes.

To improve RCNN's speed, the authors proposed Fast-RCNN. Unlike RCNN's individual feature extraction for each candidate region, Fast-RCNN uses the entire image as input for feature extraction, significantly reducing computational load. To further reduce computation, researchers introduced the Region Proposal Network (RPN). RPN is a fully convolutional network that integrates candidate box extraction and target prediction, further reducing computational requirements while improving network performance.

2.1.2 SSD Algorithm

SSD is a single-stage object detection algorithm known for its speed. The network consists of a VGG16 backbone and feature extraction layers. SSD employs two key techniques: First, multi-scale prediction detects targets from feature maps at different scales within multi-layer networks, handling various object sizes and improving detection rates. Second, SSD predicts multiple bounding boxes with different aspect ratios, storing them and adjusting through confidence scores and non-maximum suppression to filter final predictions and determine categories. The SSD network is relatively simple, integrating object detection into a single network to accelerate speed, while multi-scale prediction enhances performance. Training requires only annotated boxes and categories

without cumbersome prior box settings, making SSD easy to train and integrate into other networks, which has led to its widespread adoption.

2.1.3 YOLO Algorithm

Redmon et al. proposed YOLO (You Only Look Once), a fast single-stage detection algorithm that maintains performance while achieving high speed. YOLO transforms object detection into a regression problem, performing direct detection through a single neural network. The algorithm first divides images into an S×S grid, with each grid predicting whether it contains an object. If an object's center falls within a grid, that grid predicts the object's category, position, and size. Each grid predicts bounding boxes and their scores, processed through non-maximum suppression to eliminate duplicate predictions. YOLO's simple structure and fast speed meet real-time detection requirements but introduce some inaccuracies. Due to limited grid numbers and prediction boxes per grid, performance suffers for dense small objects.

Considering both performance and speed, subsequent experiments adopt the YOLOv5 object detection algorithm, with its network structure shown in Figure 1 [Figure 1: see original paper].

2.2 YOLOv5 Network Overview

At the input stage, YOLOv5 employs Mosaic data augmentation, which combines four images through random scaling, cropping, and arrangement. This enriches the dataset, and random scaling of small targets improves detection performance for small objects while enhancing network robustness. This four-image combination allows the model to simultaneously input data from four images during training, achieving better results even on a single GPU.

The YOLOv5 Backbone adopts the Focus module, which uses a slicing structure to reduce image size and increase channel numbers. The slicing operation is shown in Figure 2 [Figure 2: see original paper]. This operation halves image dimensions while quadrupling channels, achieving downsampling without information loss. Another module is the CSP structure, which addresses computational load issues, optimizes network computation, and reduces requirements while maintaining accuracy.

The overall structure resembles FPN+PAN architecture, as shown in Figure 3 [Figure 3: see original paper]. In object detection, deep feature maps carry strong semantic information but weak positional information, while shallow feature maps carry strong positional information but weak semantic information. FPN transfers deep semantic information to shallow layers, while PAN transfers shallow positional information to deep layers, thereby enhancing semantic expression and object localization across multiple scales and further improving feature extraction ability.

2.3 Algorithm Input and Output

Object detection algorithms generally receive RGB images as input. They must predict object categories and positions, so output typically consists of category information and bounding box representations. Categories are determined by confidence scores, with the algorithm selecting the highest-confidence category as the final prediction. Networks set confidence thresholds, where objects exceeding the threshold are positive samples and those below are negative. Bounding box positions are generally described using vertex or center coordinates plus width and height.

3. Generation of Adverse Lighting Images

Image input from cameras is susceptible to lighting and weather effects, causing varying distortion under adverse lighting. Obtaining adverse lighting images in real environments is difficult, and original-adverse image pairs cannot be acquired. However, adverse lighting images can be generated through image processing algorithms, with noise intensity quantifiable through parameters. Therefore, we can obtain adverse lighting condition images through existing algorithms.

Strong light and low light image generation: Computers commonly use the RGB color model. By converting RGB to HSV (Hue, Saturation, Value) color space, we can control lighting intensity by adjusting the brightness component before converting back to RGB mode to obtain datasets with arbitrary lighting intensity.

Haze image generation based on optical models: The mathematical relationship between hazy and original images can be expressed as:

$$I(x) = J(x)t(x) + L(1 - t(x))$$

where I(x) is the hazy image, x is the pixel coordinate, J(x) is the original image, L is the global atmospheric light component, and t(x) is transmittance. In hazy scenes, fog weakens object reflected energy J(x)t(x) while atmospheric scattering L(1-t(x)) increases, reducing image saturation. Thus, hazy images can be considered as formed by both object reflected light and atmospheric scattered light.

Assuming fog is generated at (x, y) with center (X, Y) and size as the fogging dimension, we introduce parameter α to control fog concentration. This strategy can quantify noise magnitude by controlling algorithm parameters, while the generated original images and corresponding adverse lighting image series can be used for comparative experiments, facilitating subsequent research.

4. Experiments and Analysis

4.1 Model Training

We trained the model using the VOC2007 dataset for 50 epochs total. Training used a batch size of 16. We first trained for 10 epochs with an initial learning rate of 0.01, then adjusted the learning rate to 0.001 for 20 epochs, and finally to 0.0001 for another 20 epochs. At this point, parameter changes were no longer significant, indicating basic model convergence.

The model's performance on the test set is shown in Table 1 .

4.2 Noise Impact Analysis

We used the trained YOLOv5 algorithm to detect objects in adverse lighting images, setting a confidence threshold of 0.3. Objects with detection confidence exceeding 0.3 were identified as positive samples. By controlling image processing algorithm parameters, we performed detection under strong light, low light, and haze conditions, statistically analyzing the detection confidence of identical objects across different noise levels. Organizing these results yielded confidence variation curves for objects under different noise conditions.

Strong light condition results are shown in Figure 5 [Figure 5: see original paper]. Overall, at low noise levels, most objects showed minimal confidence changes. Between noise levels 1-10, objects 3, 5, and 8 experienced significant confidence drops below the threshold, while other objects remained relatively stable. As noise levels continued increasing, objects 1 and 2 also showed declining confidence.

As shown in Figure 6 [Figure 6: see original paper], low light conditions produced more stable confidence compared to strong light. Most objects' confidence was nearly unaffected, and due to weaker reflections, some objects even showed increased confidence.

Under haze conditions, most objects were significantly affected, though a few experienced milder impact due to lower fog concentration, as shown in Figure 7 [Figure 7: see original paper].

Through feature map visualization, we compared shallow feature maps of original images with those under different noise conditions, with results shown in Figure 8 [Figure 8: see original paper].

Noise introduction reduces image contrast, blurs object features, and increases algorithmic difficulty in recognizing object characteristics, thereby affecting detection results. Among the three conditions, low light minimally impacts features, while strong light and haze have greater impacts. Due to varying object characteristics, impact degrees differ. Small and occluded objects contain less recognizable information, causing their confidence to quickly fall below thresholds under noise influence. Additionally, objects with low background contrast are prone to misjudgment under noise.

By quantifying noise magnitude and controlling noise levels, this experiment statistically analyzed detection confidence across different noise levels under three scenarios: strong light, low light, and haze. As image noise increases, misjudgment probability in feature maps rises, affecting detection results. Under strong light, most objects' confidence is affected after noise level 10. Under low light, confidence changes minimally, with only a few objects changing after level 11. Under haze, most objects show significant confidence drops after level 8. Moreover, noise impacts objects unevenly due to varying intrinsic factors.

However, it should be noted that the analyzed data was simulated. Although verified with real-world scene data, discrepancies between simulated and real data remain. Due to insufficient computational resources, the trained network's detection capability is also weaker than networks trained on servers.

Note: Figure translations are in progress. See original paper for figures.