Proceedings of the 5th International Conference on Computing and Data Science
DOI: 10.54254/2755-2721/20/20231097

Research on image recognition based on the yolov3 model in an extreme environment

Haiyang Jiang

Xidian University, 710100, China

673542471@qq.com

Abstract. In recent years, the human demand for the exploration and application of extreme environment is increasing. However, the light, climate, object motion and other factors in extreme environments have great uncertainty and complexity, which makes the image recognition technology face great challenges. This study aims to investigate the image recognition techniques based on the yolov3 (You Only Look Once Version 3) model in extreme environments. For the problem of image recognition in extreme environments, this study compared the identification gap between the initial data set in the yolov3 model, and optimized the yolov3 model (The main way is to prune and quantify, and fine-tune the algorithm) to improve its accuracy and stability in extreme environments. In this study, the feasibility of yolov3 model for image recognition in extreme environment was verified by comparing the performance of yolov3 model before and after optimization. The experimental results show that the image recognition technology based on yolov3 model proposed in this study has high accuracy and stability in extreme environments.

Keywords: extreme environment, image recognition, yolov3 model, deep learning, convolution neural network

1. Introduction

Since the 21st century, the human demand for the exploration and application of extreme environment is increasing. Extreme environment refers to those environments where conventional equipment is difficult to work and conventional technology is difficult to apply, such as polar region, deep sea, high altitude, battlefield, etc. In these environments, human beings are facing great challenges and risks, and the application of image recognition technology can provide more convenient and efficient tools and support for human beings in these environments.

Image recognition technology refers to the processing and analysis of images by computer, so as to realize the recognition and understanding of the target objects, scenes and information in the image. With the continuous improvement and application of deep learning technology, image recognition technology has made significant progress in the fields of target detection, face recognition, vehicle recognition, security monitoring and so on, and to a large extent has improve the way people produce and live [1].

However, in extreme environments, traditional image recognition techniques are often not competent. This is because the light, climate, object motion and other factors in extreme environments have great uncertainty and complexity, which makes the image recognition technology face great

DOI: 10.54254/2755-2721/20/20231097

challenges. How to improve the application ability of the image recognition technology in the extreme environment has become one of the important directions of the current image recognition technology research.

This paper aims to explore the application of image recognition based on yolov3(You Only Look Once Version 3) model in extreme environment, and improve the recognition performance of the model in extreme environment through the improvement and optimization of yolov3 model. This study aims to improve the application ability of image recognition techniques in extreme environments and promote human exploration and work in extreme environments, which has important theoretical significance and application values.

2. Research status

At present, deep learning-based object detection technology has made great progress in the field of image recognition. Among them, YOLOv3 is a popular target detection algorithm. YOLOv3 Model is highly efficient, accurate and real-time, so it is widely used in many application scenarios. However, in extreme environments, such as high altitude, low temperature, high humidity, so further study and optimization are needed.

Domestic and foreign studies show that there are many feasible solutions to the target detection problem in extreme environments. For example, some researchers use multiscale image enhancement techniques to increase the efficiency and accuracy of models. Moreover, some researchers use data augmentation techniques to increase the diversity of the datasets and thus improve the generalization ability of the model. In addition, some researchers improve the performance of the model by improving the model's architecture, such as increasing the attention mechanism and introducing multi-task learning [2].

In the past exploration, there were various technologies, algorithms, and processing ideas for studying this problem. Traditional image denoising methods mainly include image denoising methods according to spatial arithmetic mean filtering, Gaussian filter, median filter, maximum filter and minimum filter. Traditional ways often lead to the loss of image details. At the same time, their denoising performance and algorithm complexity are also unsatisfactory. Due to the significant improvement in hardware performance, the revival of artificial neural network appeared again. Many scholars have used it to image denoising, such as MLP (Multi-layer Perceptron), CNN (Convolutional Neural Network), etc., and achieved excellent results. The application of deep-learning in image denoising also demonstrates its great potential [3].

Furthermore, the application of deep learning in image denoising is a frontier direction. Despite the rapid improvement of deep-learning algorithms, they generally face huge challenges in image denoising. First, deep-learning is not suitable for various noise types and intensities, and second, an algorithm usually only applies to one situation and does not have universality. Considering these issues, more noise types and intensities can be added to the training sets, and the model used during image denoising can be optimized to improve the overall recognition ability of the model under different noise types and intensities.

3. Experimental method

3.1. Structure of the Yolov3 model

Yolov3 is an efficient target detection algorithm, which adopts a new structure and can maintain high accuracy while realizing real-time detection. The structure of Yolov3 is composed of multiple convolution layers and pooling layers, which mainly includes three parts: feature extraction network, detection network and post-processing network [4].

(1). Feature extraction network

Yolov3 The feature extraction network used is Darknet-53, which is a 53-layer convolutional neural network that uses the residual structure and is able to extract richer feature information. The feature extraction network uses the input image for multiple convolution and pooling operations to obtain a series of feature maps, which contains the position and feature information of the object.

(2). Detection network

The detection network of YOLOv3 contains three detection layers of different scales, corresponding to three feature maps of different sizes. Each detection layer will convolve the feature graph to obtain the detection results. Each detection layer outputs three anchor frames of different sizes to detect objects of different sizes. In each anchor box, the object category, location, and confidence are predicted.

(3). Post-processing network

The post-processing network of Yolov3 was used to screen and adjust the test results. First, all the test results will be ranked according to the confidence level, and the results with low confidence will be discarded. Then, the non-maximum suppression (NMS) algorithm is used to merge the final test results. Figure 1 shows the structure of the yolov3 model .

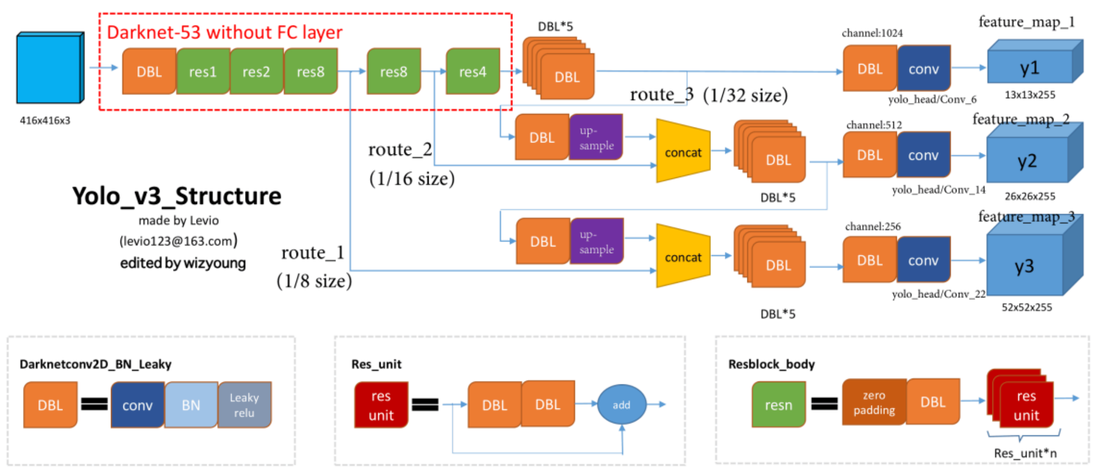

Figure 1. Structural diagram of the yolov3 model [5].

3.2. *Principles of the Yolov3 model*

The working principles of yolov3 are provided as follows [6]:

(1). The images are divided into SS grids, and each grid is responsible for predicting the presence of a target in that grid and the location and category of the target.

(2). For each grid, predict B bounding boxes (bounding box), and each bounding box contains 5 values: x, y, w, h, and confidence. Where x and y represent the offset of the central coordinates of the bounding box relative to the current grid, w and h represent the width and height of the bounding box and confidence indicates the confidence of the target in the bounding box.

(3). For each bounding box, the probability of C categories is predicted, the probability that the target in that bounding box belongs to each category.

(4). The final prediction result is the comprehensive score of the confidence and category probability of all bounding boxes in all grids, and the bounding box with the highest score is the location and category of the target.

3.3. *Experimental parameters*

Hyperparameters refer to the parameters that need to be set manually in machine learning algorithms, which are not automatically learned by training data. The setting of hyperparameters directly affects the algorithm performance and training speed. The main hyperparameters used in this experiment are

DOI: 10.54254/2755-2721/20/20231097

as follows:

(1) Learning rate: control the step size of the weight in each iteration. The large learning rate will cause the model to converge, while the small learning rate will lead to the slow model convergence rate. The initial learning rate used in this experiment was 0.001.

(2) Regularization coefficient: used to control the complexity of the model and prevent overfitting. The larger the regularization coefficient, the lower the complexity of the model, but may lead to underfitting.

(3) Batch size: refers to the number of samples used in each iterative training process. Too small batch size will lead to the slow model convergence rate, and too large batch size will make the model difficult to converge. The batch size used in this experiment is 64.

(4) Activation function: The activation function determines the output of the neurons. Different activation functions have an impact on the performance of the model. The activation function used in this experiment is the ReLU(Rectified Linear Unit) function, and the formula is as follows.

$$f(x) = max(0, x) \quad (1)$$

(5) Loss function: a function used to assess the difference between the model prediction result and the true value. In machine learning, the goal of the model is to minimize the loss function to improve the accuracy and generalization ability of the model, and the loss function used in this experiment is the cross-entropy loss function, as formulated as follows [7].

$$L = -\sum_{i=1}^{\infty} y_i log(\hat{y_i}) + (1 - y_i) log(1 - \hat{y_i}) \quad (2)$$

3.4. Experimental Methods

In this study, the experimental method is mainly used, first, the public data set (COCO data set) on the network, and then the images are processed in batches (add noise, simulating extreme weather), image preprocessing in the model, and image detection, recording the accuracy of recognition and other parameters. Based on this, the algorithm is optimized to make it more suitable for target detection in extreme environments. Here are the specific experimental steps:

(1). Preparation of datasets: Prepare datasets and add noise to the original datasets to represent the simulation of extreme weather, including extreme environments, such as bad weather, insufficient light, and high altitude, to ensure the diversity and authenticity of the datasets. In this experiment we selected a part of the COCO(Common Objects in Context) datasets for detection, as shown in Figure 2.

(2). Model selection: In this experiment, we chose the yolov3 model with high performance in the field of target detection as the basic model of this study, and tested it with GPU in the pytorch framework.

(3). Data preprocessing: Conduct the preprocessing of the collected image data, including image enhancement, image cropping, data annotation and other operations [8].

(4). Model training: First, the original data set is used to train our yolov3 model to obtain the accuracy of the yolov3 model under original datasets.

(5). Model test: Apply the trained model to the image recognition task in the extreme environment, and the performance of the model is evaluated by calculating the accuracy and recall rate of the model on the test set.

(6). Model optimization: optimize and fine-tune the yolov3 model on the basis of the above experiments, so that the efficiency of model identification can be improved without affecting the identification accuracy, and the above experiments are carried out again. The main way of model optimization in this experiment is to prune and quantify, and fine-tune the algorithm [9].

(7). Analysis of the experimental results: the experimental results were analyzed to explore the image recognition performance of the yolov3 model in the extreme environment, and the influence of different factors on the model performance.

DOI: 10.54254/2755-2721/20/20231097

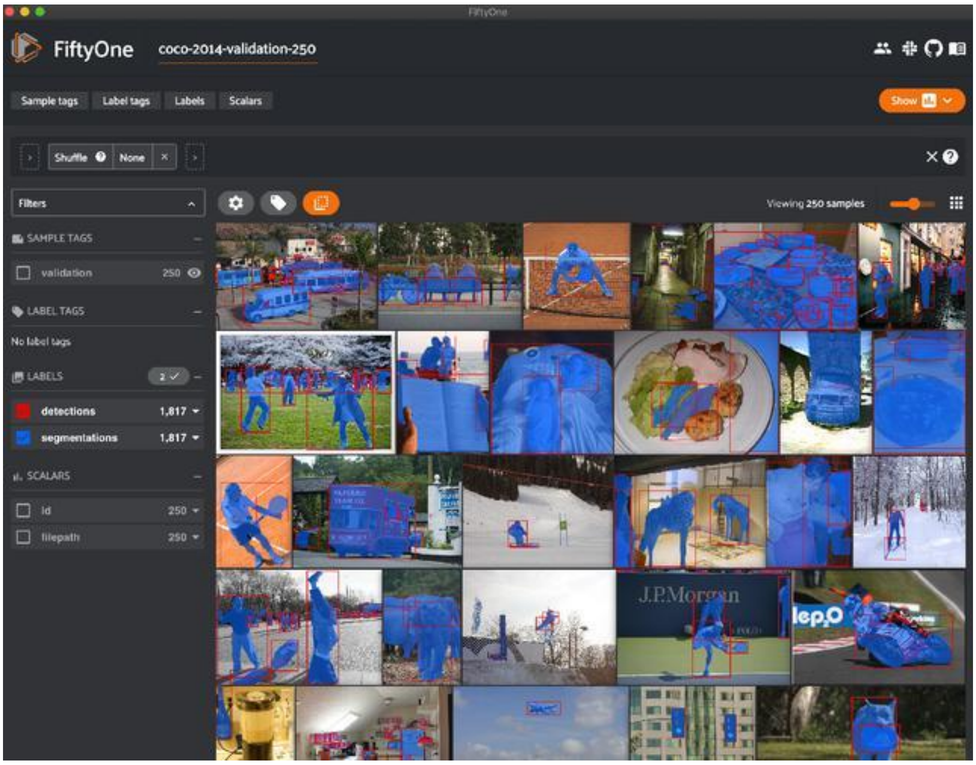

Figure 2. The COCO datasets [5].

4. Experimental results and data analysis

In this experiment, a total of four training sessions were performed. The initial data set is trained and the noise-added data set is trained before and after model optimization. The parameters used to evaluate the recognition ability of yolov3 model mainly include accuracy, recall, F1 value and processing time. The following table shows the results obtained in the initial experiment.

Table 1. Experimental data before model optimization.

	Precision	Recall	F1 value	Handling time
Original datasets	74.3%	71.8%	73.0%	5.2min
Noisy datasets	65.3%	72.6%	68.7%	6.8min

The following table shows the experimental results obtained after the model optimization.

Table 2. Experimental data after model optimization.

	Precision	Recall	F1 value	Handling time
Original datasets	71.1%	72.3%	71.6%	6.1min
Noisy datasets	68.6%	73.1%	70.9%	7.9min

Through the training results of the original datasets and the noise datasets, it is not difficult to realize that the accuracy of noise datasets is reduced after the initial training through the original data set, but the reduction is not very large. However, after model optimization, the difference between accuracy values is significantly reduced, and the representative model is significantly more suitable for target detection in extreme environments. At the same time, the training time gap before and after the model optimization is small, which meets the actual requirements.

DOI: 10.54254/2755-2721/20/20231097

5. Conclusion

The present study aims to explore the image recognition technology according to the yolov3 model in extreme environments. We employed datasets with multiple target categories and performed extensive experiments and analysis. Through experiments, we obtained two methods to improve the image recognition accuracy of yolov3 model in extreme environment:

A lot of initial training before image recognition, so that the yolov3 model can identify a large number of objects. On this basis for image recognition in extreme environment, the accuracy of image recognition will be higher.

Model optimization, including but not limited to pruning, quantification, algorithm fine-tuning mentioned in the text, can also be a lot of training to obtain appropriate hyperparameters, or model structure optimization, will have a significant role in image recognition in extreme environments.

References

[1] Redmon J , Farhadi A .YOLOv3: An Incremental Improvement[J].arXiv e-prints, 2018.

[2] Zhang, H., Cao, Z., Liu, Y., & Zhang, L.(2018).Learning deep cnn denoiser prior for image restoration.In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (pp.2808-2817).

[3] Cai R .Research Progress in Image Denoising Algorithms Based on Deep Learning[J].Journal of Physics Conference Series, 2019, 1345:042055.

[4] Pang T , Zheng H , Quan Y , et al.Recorrupted-to-Recorrupted: Unsupervised Deep Learning for Image Denoising[C]// Computer Vision and Pattern Recognition.IEEE, 2021.

[5] CDSN. (2018) Yolo v3 of yolo series. https://docs.voxel51.com/integrations/coco.html/

[6] Liu J ,Liu X ,Shao Z , et al.Research on blind restoration of noisy blurred image based on deep l earning[C]// ICASIT 2020: 2020 International Conference on Aviation Safety and Informati on Technology.2020.

[7] Gai Shan, Bao Hongyun. Deep learning-based high-noise image denoising algorithm [J]. Journal of Automation, 2020,46 (12): 9.

[8] Li Chuanpeng. Image denoising study based on deep convolutional neural networks [J]. Computer Engineering, 2017, (3): 253-260.

[9] Li Songshun, Zhou Junwei, Du Zhenhua, et al. A model optimization algorithm based on the target detection YOLOv3 of deep learning:, CN112001477A [P].2020.