Based on Weak Light YOLOv3 Multi-Target Detection

Ding Aodi
Faculty of Applied Mathematics and Computer Sciences
Belarusian State University
Minsk, Belarus
aodiding541@gmail.com

Pavel Lukashevich
United Institute of Informatics Problems,
National Academy of Sciences
Minsk, Belarus
ORCID: 0000-0002-8544-8554

***Abstract*—Inside the real life scenarios, the YOLOv3 target detection model has achieved good results on many benchmark datasets.the light illumination conditions are poor in many scenarios, such as night, indoor, foggy weather, in dark conditions is still a huge challenge, so in This environment first use the filter to process the image, due to the filter to process high-resolution images is very costly computer resources, so I will use the filter alone to process the filter parameters obtained from high-resolution images transplanted to the original resolution of the image of the model for this experiment, in this experiment choose the detector YOLOv3 as the detection network, YOLOv3 based on the idea of residual network optimization Network multilayer structure can further improve the detection accuracy, especially for small targets, in this yolov3 strengthened the discovery of potentially beneficial information in the image, so the image can be detected in low light with the support of this model framework.**

Keywords—yolov3, filter, target detection, residual network

I. INTRODUCTION

The rapid rise of artificial intelligence in recent years, especially in the field of image recognition, automatic driving, medical, military field of rapid development, scientific researchers and scholars with the upgrading of the model network and optimization of the image recognition accuracy, there is a core component has not been changed convolutional neural network, in the majority of the image recognition in the convolutional neural network has been active in the forefront of scientific research, in the use of the YOLOv3 network is also The most basic convolutional neural network is used, YOLOv3 is a neural network model for the task of target detection, Yolov3 uses more convolutional layers and larger input sizes, which makes the network deeper and wider, and is able to detect smaller objects.Yolov3 also uses three feature maps at different scales to detect objects of different sizes, which improves the accuracy of detection and the recall rate. Recall[10]. In addition, Yolov3 introduces a new technique called "Bag of Freebies" that improves model performance through data augmentation, improved training strategies, and better network initialization. Yolov3 has better performance and higher accuracy than Yolov2, especially in detecting small objects[9].

When the target detection model of Yolov3 achieved good results on many benchmark datasets, but detecting targets in low light conditions is still a great challenge, weak light contains images obtained in dark environments, images after camera exposure, and images in dim conditions have a significant reduction in recognition accuracy using YOLOv3. Although many approaches have been proposed to address the robustness problem in dark scenes, for example, many shimmer enhancement models have been proposed to recover image details and reduce the effects of poor lighting conditions[11]. However, the complex structure of the shimmer enhancement models is not conducive to the real-time performance of the detector after image enhancement. Most of these methods cannot be trained end-to-end with the detector and require supervised learning of pairs of shimmering and normal images, resulting in a significant increase in computational resources, which degrades the performance of yolov3, so to address such problems I found that the degradation of image quality severely affects the performance of the YOLOv3 detector, and that high-quality images help the model to improve its better High quality images can help to improve the accuracy of the model, so I need to use a suitable filter to improve the clarity of the object lines in the image before it enters the YOLOv3 model, and use the filter to enhance the detail performance of the image, which introduces the design of the filter can be categorized into pixel-level filters, particle removal filters, and sharpening. Pixel level filters pixel level filters map an input pixel value to an output pixel value which makes the function microscopic with respect to the input image and the parameters, granular object removal filters remove the impurity information from the image, but using these filters can help to enhance the saliency of the information in the image, filters combined with the yolov3 model can improve the target detection accuracy in low light.

II. RELATED WORK

A. Data Preparation

The purpose of this experiment is mainly for multi-target detection in low light, but it is difficult to find a lot of pictures in low light, so I will make changes to the photos in the data set, I am using a software tool to reduce the saturation or chroma of the image, or use blurring tools to reduce the clarity of the picture finally collected 20,000 pictures of the data under different optical fibers.

The main training scheme is still using model data. The backbone network used is Darknet-53. during the training process, I am scaling the training dataset with resizing the image to (32N, 32N), using data enhancement methods such as image flipping, cropping and transforming. The model was trained using the Adam optimizer and trained for 200 epochs. the starting learning rate was , and the batch size size was 6. The model predicted three bounding boxes on three different scales, and three anchor points on each scale.

B. Model Functions

This experiment uses a total of two models combined to recognize the results of multi-target detection.

The first model: filter parameter prediction model, this model is to get the filter parameters, because for different image processing need to use different filter parameters.

The second model: YOLOv3 model, this model is a mature multi-target detection model, YOLOv3 model compared to the previous version, increased the Resnet idea can be stacked more layers to extract features, multiple scare detection of different sizes of objects, 9 kinds of a priori

frames are designed to optimize the accuracy of the detection of small targets.

III. Model Structure And Principles

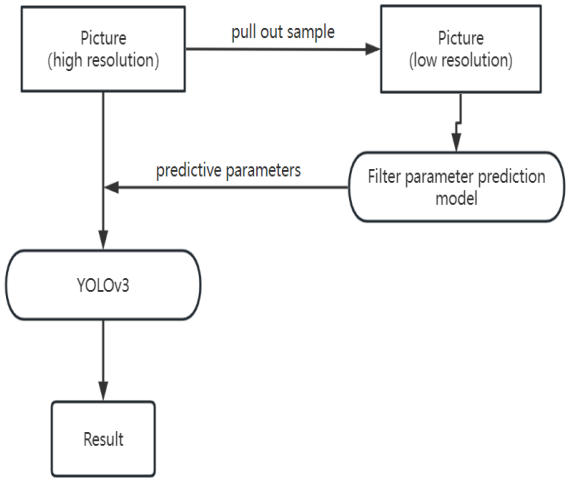

Fig. 1. Weak light model structure graph

From the Figure 1 we can see the image first after downsampling changes into the filter prediction model, the filter model to get the prediction results through the original image to do adjustments, which will make the dataset image details become clearer, and finally after the YOLOv3 model to recognize the target detection to get the results.

A. *Sharpening Filter*

The first-order derivative is not zero when the image gray level changes, the second-order derivative is negative when the gray value starts to decrease, and positive when the gray value stops decreasing. The purpose of sharpening is to highlight the excessive portion of the gray scale, i.e., the image edges[3]. In order to highlight the image edge, the part of the image edge with large gray value should increase the gray value to make it larger, and the part with small gray value should decrease the gray value to make it smaller, sharpening filter highlights the edge information of the object at the same time of denoising, keeps the image edge information unchanged or enhances the edge information. The essence of image sharpening is to enhance the high frequency component of the original image, the image sharpening filter is a high pass filter, the edges and contours are generally located in the gray level of the sudden change, so you can use the gray level difference to extract the image edges and contours[13]. Since contours and edges often have arbitrary directions in a picture, and the difference operation has directionality, if the direction of the difference operation is not selected appropriately, the edges and contours that are inconsistent with the direction of the difference will not be detected. Image sharpening has the ability to detect edges and contours in any direction. Image sharpening brings out the details of an image. Like the unsharpened mask technique, the sharpening process can be described as follows [1].

$$F(x,\lambda) = I(x) + \lambda(I(x) - \mathrm{Gau}(I(x))) \qquad (1)$$

Formula: $I(x)$ is the input image, $\mathrm{Gau}(I(x))$ is the Gaussian filter, and λ is the positive scale factor. This sharpening operation is differentiable for both x and λ. The degree of sharpening can be adjusted by optimizing λ to get better target detection performance[8].

B. *Granular Object Removal Filters:*

Dark channel fogging transmittance map than other algorithms are fine, if appropriate to reduce the accuracy of a little bit, the fogging effect should not be too different in theory, so I thought of a way, that is, the transmittance of the original map is not the time to get the original map, but the original map first downsampling, such as shrinking to the original map 1/4, calculate the transmittance of a small map, and then through the interpolation of the way After that, we can get the approximate transmittance of the original image by interpolation, and then we should be able to get the effect[12]. After practice, this approach greatly improves the speed of implementation, and the effect and the original program is basically the same, for 1024 * 768 images only need about 30ms, if further scaling to take 1/9, it only takes about 20ms, can fully meet the industrial real-time requirements of high occasions.

The removal granularity filter equation is shown below w is used to control the degree of removal of particulate matter, atmospheric light A, t(x) is the medium penetration function, and i is the atmospheric scattering model [2].

$$t(x,\omega) = 1 - \mathrm{wmin}_C\left(\min_{y\in\Omega(x)} \frac{I^C(y)}{A^C}\right) \qquad (2)$$

The above function is differentiable, so training ω by backpropagation allows the granular object removal filter to provide a good basis for subsequent target detection.

C. *Filter Parameter Prediction Model*

The selection of filters should follow microscopicity in order to allow training the network by back propagation. However, in order to calculate the filter parameters corresponding to the picture, let all the pictures into the convolutional layer to calculate this will greatly consume the computer computing space and computing time, in order to optimize such a problem, I put all the picture data using the downsampling method to let the resolution of the picture reduced, the low-resolution picture into the filter parameters prediction model, so as to reduce the running space of the computer and speed up the model This reduces the running space of the computer and speeds up the model [7].

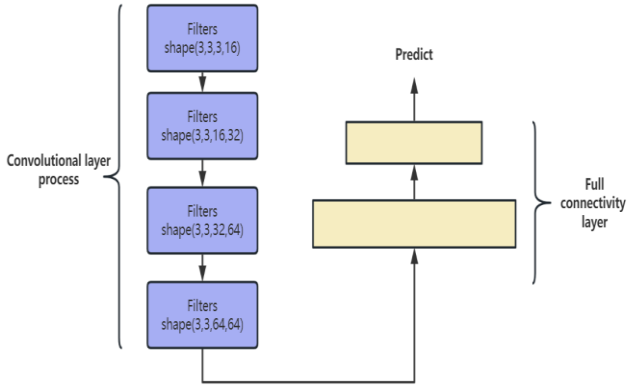

Fig. 2. Convolutional structure graph

From Figure 2 we can understand filter parameter prediction model using the size length 3 width 3 convolutional kernel for four convolutional feature extraction, four convolutional channel output were 16, 32, 32, 64, and then after two fully connected network to predict the completion of the filter's parameter values, through the use of predicted completion of the filter's parameter values with the use of the filter in the original image for filtering will be obtained in the details of the lines of the picture is more pronounced, and then for the recognition of the YOLOv3 target detection [4].

D. YOLOv3 Model

YOLO v3 adopts the up-sampling and fusion practice, From Figure 3 we can understand three scales of large, medium, and small scale maps, respectively, and doing the detection independently on the fused feature maps of multiple scales, which eventually improves the detection effect of small targets significantly through many experiments [6].
In YOLOv3, the number of a priori frames is changed to 9, and the selection of a priori frames is generated by K-means clustering algorithm. Three prior frames are assigned under each scale image. Each cell outputs (1+4+C)*3 values, which is the depth of the final output feature tensor at each scale. Although YOLO v3 predicts 3 bounding boxes per 1 cell, the number of bounding boxes is much more than the previous version because YOLO v3 uses multi-scale feature fusion.

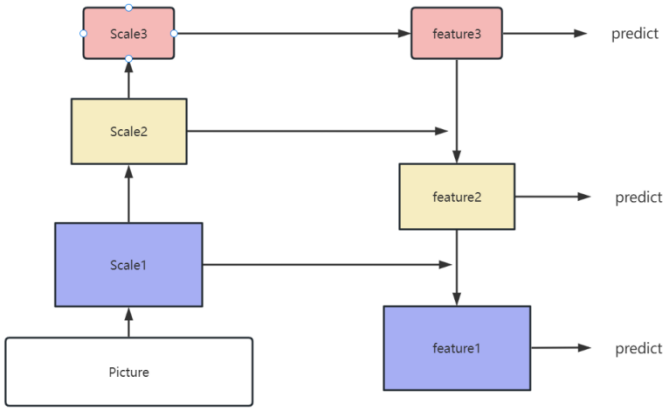

Fig. 3. YOLOv3 model scale graph

Model used this time is composed of 53 convolutional layers, including 1×1 and 3×3 convolutional layers, convolution saves time and effort and is fast and effective, and is most effective for analyzing object features [15]. After each convolutional layer contains a batch normalization layer and a Leaky ReLU, the purpose of adding these two parts is to prevent overfitting, not set up a fully connected layer is because it can be adapted to all sizes of the input image, there is no pooling layer, the convolutional layer by controlling the convolutional layer conv's stride stride to achieve the effect of downsampling so that the model can be made faster, the use of three kinds of sizes Different sensory fields are used so as to detect small targets more accurately.

From the Figure 4 in the weak light line, the lines and contours of the picture are not very clear, which makes the picture in the use of convolution to extract features can not be used to extract the picture details and small target information features completely [5].We can know from Figure 5 the picture after the filter adjusted, the outline of the picture is

Fig. 4. Image without filter model optimization graph

Fig. 5. Image filter model optimization graph

more clear, especially in the small target of the lines of the contour is more obvious, when the YOLOv3 extraction of the small target will be more easily and will not affect the speed of the operation of the YOLOv3 [14].

IV. ANALYSIS OF MODEL RESULTS

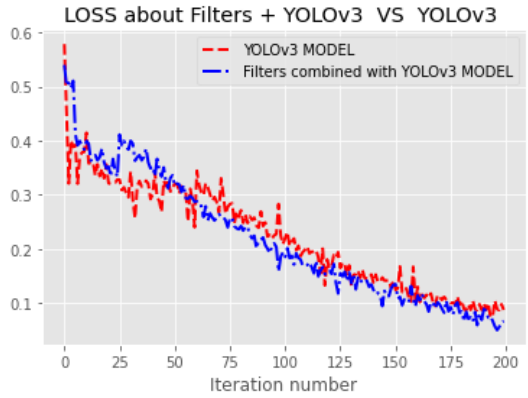

Fig. 6 Comparative losses

We can see from Figure 6 the blue model is the model loss results of this experiment, the red is just used YOLOv3 model loss results, from the picture we can understand that the two models have been trained after 200 iterations, after 200 times the blue model loss is less than the red model surface through the use of filter combined with the YOLOv3 model so that the loss reaches a lower.

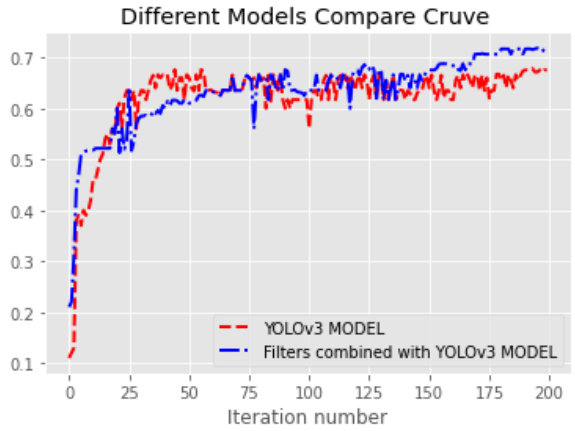

Fig. 7. Comparative accuracy

Through the Figure 7 two models accuracy analysis can be seen in blue is the filter combined with the YOLOv3 model after 200 training 150 times began to steadily improve the accuracy rate, and finally after 200 training to achieve an accuracy rate of 71%, just with the YOLOv3 model after 200 training to achieve an accuracy rate of 66%, through the two models compared to these accuracies can be concluded that the use of the filter YOLOv3 model in the multi-target detection will increase the rate of correctness of 5%, which can be effective in helping to use the effect of the filter in the low-light will allow the accuracy rate to achieve a better result.

V. Conclusions

Inside the real life scene, many scenes under the light illumination conditions are relatively poor, such as night, indoor, foggy days, etc., in order to improve the picture is modeled to capture with more feature information, the filter can enhance the picture edge branch can enhance the texture of the components and enhance the details of the enhanced image. Based on these application scenarios this experimental study in the weak light using a combination of filter and YOLOv3 method compared to the normal target detection model YOLOv3, after 200 times of training YOLOv3 model accuracy is only 66%, the filter and YOLOv3 combination model to the data photo downsampling to get a smaller picture, into the filter parameter prediction model, to get the filter parameter Post-processing the image and then put it into the YOLOv3 model will make the target detection accuracy in weak lighting reach 71%, which is a 5% increase in accuracy without affecting the speed of the model.

Acknowledgment

First of all, I am very grateful to my Александр Михайлович tutor, who guided me with solid theoretical foundations and a wealth of professional knowledge, tutor guided in the direction of my guidance in the field of multi-target image recognition, and when I had difficulties he was always kindly pointed out to me my mistakes and gave me guidance!

References

[1] Hudnell M, Price T, Frahm J M. Robust aleatoric modeling for future vehicle localization[C]// 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW). Long Beach: IEEE, 2019: 2944. DOI: 10.1109/CVPRW.2019.00355

[2] Sadeghian A, Kosaraju V, Sadeghian A, et al. SoPhie: an attentive GAN for predicting paths compliant to social and physical constraints[C]// 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Long Beach: IEEE, 2019: 1349. DOI: 10.1109/CVPR.2019.00144

[3] Choi W, Savarese S. A unified framework for multi-target tracking and collective activity recognition[C]// European Conference on Computer Vision. Berlin: Springer, 2012: 215. DOI: 10.1007/978-3-642-33765-9_16

[4] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In CVPR, 2016. 1

[5] Zhang Hongyi, Cisse Moustapha, N. Dauphin Yann, and David Lopez-Paz. mixup: Beyond empirical risk minimization. ICLR, 2018. 3

[6] Xin Huang, Xinxin Wang, Wenyu Lv, Xiaying Bai, Xiang Long, Kaipeng Deng, Qingqing Dang, Shumin Han, Qiwen Liu, Xiaoguang Hu, et al. Pp-yolov2: A practical object detector. arXiv preprint arXiv:2104.10419, 2021. 3, 6

[7] Kang Kim and Hee Seok Lee. Probabilistic anchor assignment with iou prediction for object detection. In ECCV, 2020. 1, 4

[8] Seung-Wook Kim, Hyong-Keun Kook, Jee-Young Sun, Mun-Cheon Kang, and Sung-Jea Ko. Parallel feature pyramid network for object detection. In ECCV, 2018. 2

[9] Hei Law and Jia Deng. Cornernet: Detecting objects as paired keypoints. In ECCV, 2018. 1, 3 [15] Mengtian Li, Yuxiong Wang, and Deva Ramanan. Towards streaming perception. In ECCV, 2020. 5, 6

[10] Tsung-Yi Lin, Priya Goyal, Ross Girshick, Kaiming He, and Piotr Dollar. Focal loss for dense object detection. In ´ ICCV, 2017. 2 [17] Tsung-Yi Lin, Michael Maire, Serge Belongie, James Hays, Pietro Perona, Deva Ramanan, Piotr Dollar, and C Lawrence ´ Zitnick. Microsoft coco: Common objects in context. In ECCV, 2014. 2

[11] H. Trimarchi, J. Barratt, D. C. Cattran, H. T. Cook, R. Coppo, M. Haas, Z.-H. Liu, I. S. Roberts, Y. Yuzawa, H. Zhang, et al., "Oxford classification of iga nephropathy 2016: an update from the iga nephropathy classification working group," Kidney international, vol. 91, no. 5, pp. 1014–1021, 2017.

[12] N. J. Vickers, "Animal communication: when i'm calling you, will you answer too?," Current biology, vol. 27, no. 14, pp. R713–R715, 2017.

[13] Z.-Q. Zhao, P. Zheng, S.-t. Xu, and X. Wu, "Object detection with deep learning: A review," IEEE transactions on neural networks and learning systems, vol. 30, no. 11, pp. 3212–3232, 2019. [17] A. Garcia-Garcia, S. Orts-Escolano, S. Oprea, V. Villena-Martinez, and J. Garcia-Rodriguez, "A review on deep learning techniques applied to semantic segmentation," arXiv preprint arXiv:1704.06857, 2017.

[14] Jifeng Dai, Yi Li, Kaiming He, and Jian Sun. R-FCN: Object detection via region-based fully convolutional networks. In Advances in Neural Information Processing Systems (NIPS), pages 379–387, 2016. 2

[15] Jia Deng, Wei Dong, Richard Socher, Li-Jia Li, Kai Li, and Li Fei-Fei. ImageNet: A large-scale hierarchical image database. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pages 248–255, 2009. 5