Proceedings of 2022 IEEE
International Conference on Mechatronics and Automation
August 7 - 10, Guilin China

RGBD image based human detection for electromechanical equipment in underground coal mine

Tao Huang, Xiaoyu Zou, Zhongbin Wang, Honglin Wu*

China University of Mining and Technology, School of Mechanical and Electrical Engineering, Jiangsu Key Laboratory of Mine Mechanical and Electrical Equipment, Xuzhou, Jiangsu, 221116, China

**Corresponding author: zouxiaoyu@cumt.edu.cn*

Qingfeng Wang

China Coal Technology Engineering Group Chongqing Research Institute, Chongqing 400039, China

***Abstract* - Human detection within the operating area of electromechanical equipment is essential to ensure safe production and avoid accidents in the underground coal mine. Low light intensity and uneven light distribution in its environment surrenders the traditional color image based methods for human detection. In this paper, we focus on accurate detection of human in the operating area of electromechanical mining equipment using RGBD image. A novel network framework for miner detection based on YOLOv3 is proposed to fuse color image and depth image with enhanced attention mechanism. In the Pre-Backbone, feature extraction of both Depth and RGB branches are developed as the preliminary feature extractor with convolutional layer and residual block. Then the Convolutional Block Attention Module (CBAM) is improved to select and fuse RGB and Depth features by defining channel weights. Finally, the features are further inputted to Post-Backbone and used for multi-scale prediction in Head. The experiments demonstrate the superiority of the proposed method over some classical methods on miner detection with different light intensities and distributions.**

Index Terms* - *Human detection, underground coal mine, RGBD image, Feature Fusion, CBAM.

I. INTRODUCTION

Electromechanical equipment in underground coal mine, such as shearer, drilling robot, belt conveyor, etc., is increasingly intelligent and complex. The operation process of the equipment is dangerous and uncertain, and thus the miners are usually prohibited from approaching near their working areas. Accurate detection of human [1] in the working area of mining electromechanical equipment, is essential for automatic object tracking [2], intrusion detection [3] of dangerous area, etc., which are basic security for preventing safety accidents and ensuring the normal operation of coal mine equipment.

With the development of computer vision technology and deep learning, more and more scholars use computer vision and deep learning methods to research on human detection of underground coal images and videos. Cai et al. [4] proposed a method to detect miners by detecting their helmets, which solved the problem that the miners could not be detected as the miners are very similar to the background in underground coal mine videos. Following that, they proposed a miner fuzzy detection method based on mixture Gaussian model, which can remove the interference of miner's lamp and detect the miner effectively even they are similar to the background [5]. Sun et al. [6] proposed the idea that pre-process face images using the 2D-wavelet-transformation-based Mallat algorithm with good detection effect for the actual situation that the in-pit personnel are wearing helmets and faces are prone to be stained during the face recognition.

Although the above methods improve the performance of human detection in underground coal mine, they do not consider the special scenario in underground coal mine. Different from the ordinary scenario, for most of the coal mines that require deep mining, the underground environment is very special and artificial light is used throughout the day. Limited by its power, artificial light has low light intensity and uneven light distribution. The color images captured by the image sensor in the underground coal mine environment have the following challenges.

1) In the low-illumination environment, the color images captured by the image sensor are not clear and the imaging effect is poor. Meanwhile, the safety clothing of underground coal mine human is usually dark in color, which is similar to the color of the coal mine environment under low illumination. The recognition of the human in the captured image is so low that it is difficult to be detected.

2) In the environment with uneven light distribution, the color images captured by the image sensor are prone to light spots and blurring. And along with the movement of people in underground coal mine, light exposure from the helmet headlamps worn on their heads can exacerbate the uneven distribution of light in the coal mine environment. Accurate detection of human in the captured color images is a challenge.

At the same time, object detection algorithms are constantly evolving, and the You Only Look Once (YOLO) algorithms [7] represented by YOLOv3 [8] are widely used in realistic scenarios [9], [10], [11]. In order to let the model focus its attention on important image information, the visual attention modules, such as Squeeze-and-Excitation (SE) block [12], CBAM [13], etc., is frequently used in object detection algorithms [14] [15]. And in saliency detection, there are many algorithms that use RGB image and Depth image as input to investigate how the two can be effectively fused. For example, Qu et al. [16] designed a new convolutional neural network (CNN) to automatically learn the interaction mechanism of RGBD salient object detection to efficiently merge low-level saliency cues to generate a master saliency map.

Due to the above problems of color images in underground coal mine, it is difficult to achieve better result

978-1-6654-0851-6/22/$31.00 ©2022 IEEE

in underground coal mine using traditional visible light-based methods. In order to solve the above problems, we use depth information for underground coal mine human detection and combine the characteristics of color image information and depth information, and propose a method to fuse color image information and depth information for underground coal mine human detection based on attention mechanism. Meanwhile, a current attention mechanism is improved to better integrate these two types of information. We adopt YOLOv3 algorithm as basis, using convolutional layer and residual block [17] for initial feature extraction of RGB image and Depth image respectively. Then attention mechanism is improved for selection and preliminary fusion of these two features, then using ResBlock [17] for deep fusion and feature extraction of the fusion-preliminary features, and finally using convolutional layer and upsampling layer for multi-scale prediction. The contributions of the present paper are as follows.

1) An accurate detection method for human is proposed in underground mine scenario, by integrating depth information with RGB image. Thus the problems of human detection with low light intensity and uneven light distribution in underground coal can be solved.

2) In the human detection algorithm for underground coal mine, an attention mechanism is improved and utilized to effectively fuse RGBD images.

The rest of this paper is structured as follows. The proposed method is introduced in the second section. Then the experiments and discussion is followed. Finally our work is concluded in the last section.

II. The Proposed Novel Network Framework for Human Detection in Underground Coal Miner

Most of the previous methods for detecting human in underground coal mine are based on color image information, i.e. RGB image. However, RGB image is not clear and has poor imaging effect in the environment of low light intensity and uneven light distribution in underground coal mine. This results in poor detection of such methods in underground coal mine environment. However, depth information is not affected by light intensity in underground coal mine. And when humans wear overalls and are close to the depth camera, their corresponding depth images have obvious features, as shown in Fig. 1. Therefore, in our method, we convert the depth information into a grayscale image, i.e. Depth image, as input.

Depth image has blurred visual effect and poor resolution compared with RGB image. So the Depth image and RGB image are used together as the input of our proposed method. We set two branches to perform preliminary feature extraction on RGB image and Depth image respectively to get RGB feature channels and Depth feature channels.

Neither using convolution operation nor using current attention mechanisms can efficiently select and fuse the two types of information for accurate detection of human, so a current attention module, CBAM [13], is improved to better select and fuse the two types of information. RGB feature channels and Depth feature channels are concatenated together as whole feature channels as the input of the improved attention module. Fusion-preliminary feature maps are obtained in the improved attention module and used for further feature extraction and human prediction. This method can achieve to focus on the more important information in color image information and depth information under various lighting intensity environments, while fusing the two kinds of information to achieve more accurate detection. We validate this approach on the basis of the YOLOv3 algorithm.

Fig. 1 Depth images of human wearing overalls in the alleyway. The human in the Depth image show distinct features due to the reflective stripes of the overalls.

YOLOv3 is a One-Stage object detection algorithm, which mainly consists of Backbone and Head. Its network structure is shown in Fig. 2. Backbone uses Darknet-53 for feature extraction and Head uses Yolo layer for multi-scale prediction. First, we use Depth image of one channel for human prediction and propose YOLOv3-Depth as shown in Fig. 3. Then we improve on YOLOv3 and propose our method.

The network structure of our proposed method mainly consists of Pre-Backbone, Post-Backbone and Head, as shown in Fig. 3. Pre-Backbone is used for feature selection, preliminary feature extraction and fusion of RGB image and Depth image. Post-Backbone is used for deep feature fusion and extraction of preliminary-fused features. Head is used for multi-scale prediction. Our improvement points are in the Input and Pre-Backbone sections. The Post-Backbone and Head sections still use the structure of YOLOv3, as shown in Fig. 2. The inputs of our proposed network structure are RGB image of three channels and Depth image of one channel. In the Pre-Backbone, we perform the convolution operation and residual connection [17] on RGB image and Depth image to obtain RGB feature channels and Depth feature channels respectively. The two are concatenated together as whole feature channels, which are used as the input of the attention module, improved CBAM.

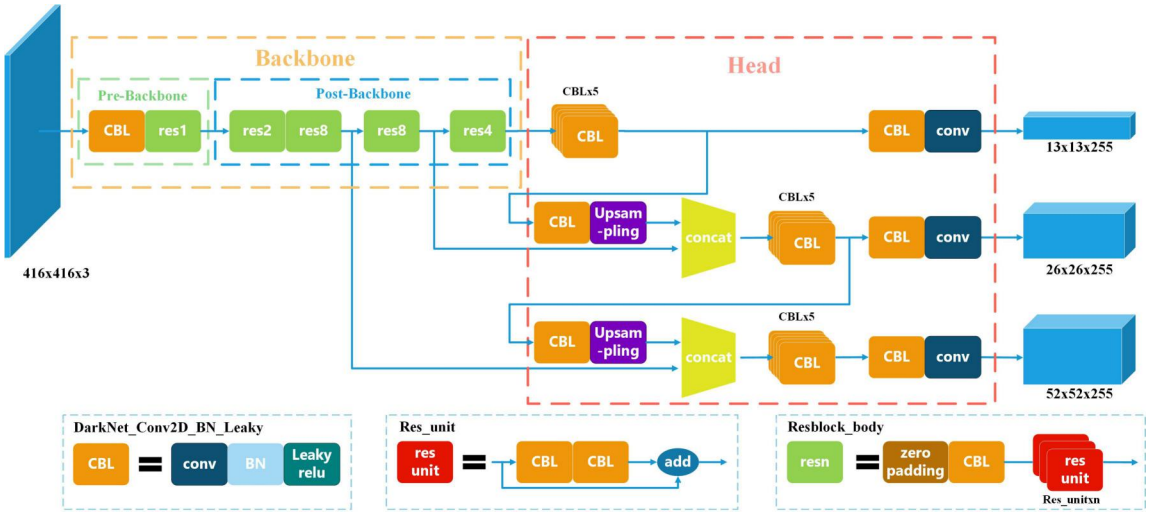


Fig. 2 The network structure of YOLOv3.

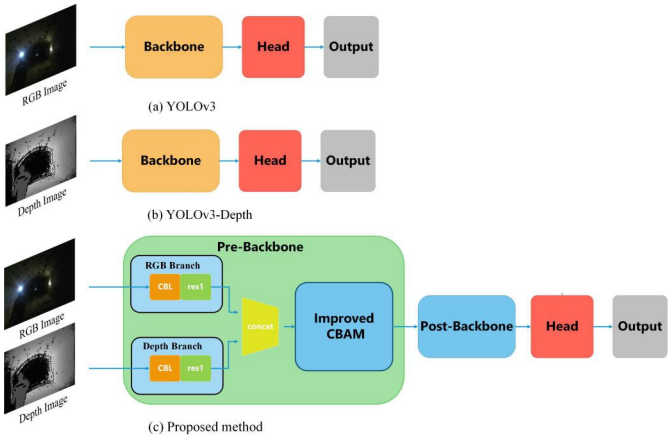


Fig. 3 The framework of YOLOv3, YOLOv3-Depth and our proposed method.

Improved CBAM consists of channel attention module and spatial attention module, as shown in Fig. 4. Only the channel attention module is improved. In the channel attention module, the whole feature channels are calculated by average-pooling layer and fully connected layer to get two weight values which correspond to the RGB feature channels and Depth feature channels, respectively, as shown in Fig. 4. After multiplying two types of feature channels with their corresponding weight value, the feature-selected channels are obtained. In the spatial attention module, the feature-selected channels are calculated by average-pooling layer and max-pooling layer to generate an efficient feature descriptor [13].And a spatial attention map is generated by applying a convolution layer on the efficient feature descriptor [13], as shown in Fig. 4. After multiplying the feature-selected channels with the spatial attention map at each pixel location, fusion-preliminary feature maps are obtained.

In the Post-Backbone, the fusion-preliminary feature maps are passed through the ResBlock [17] for deep feature fusion and extraction to obtain final feature maps. Finally, in the Head section, the final feature maps are convolved and upsampled for multi-scale prediction.

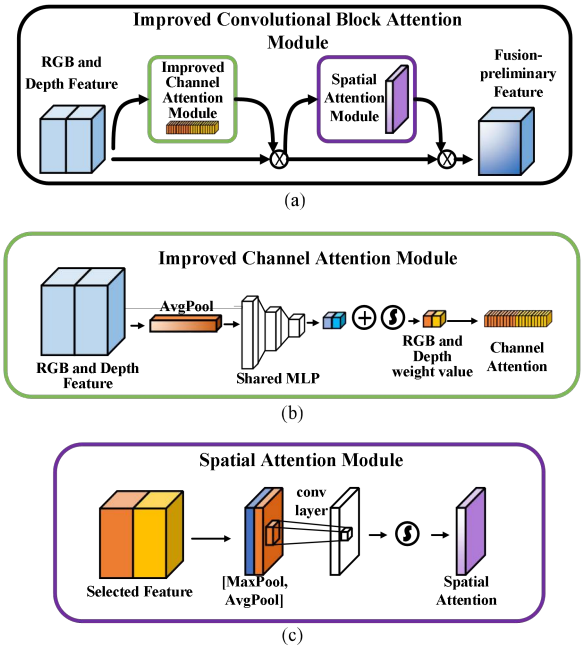


Fig. 4 Improved convolutional block attention module.

III. Experiments and Discussion

A. Datasets and Experiment Setup

For human detection in underground coal mine, one of the problems is the shortage of data sets. There is no open source RGBD dataset [18] for underground coal mine human detection available in the Internet. For this reason, we created a RGBD dataset, CUMT_RGBD_Miner. The train set of CUMT_RGBD_Miner contains 4614 RGBD images captured in the lab with Microsoft Kinect v2. To prevent overfitting due to the homogeneity of the lab scene, we enhanced 2984 images by conventional image processing algorithms using NYU Depth Dataset V2 [19] and KITTI data set [20] as the background sets. The test set consists of 2343 images captured with Microsoft Kinect v2 in a simulated alleyway.

The simulated alleyway have low light intensity, uneven light distribution, and is accompanied by miner's headlamp lighting. As shown in Fig. 5, it is difficult to perform human detection through RGB image alone. At the same time, the activity range of miners in the collected data is within the range of 0-8 m from the depth camera, the human in the Depth image have obvious characteristics. In this paper, we will conduct experiments on CUMT_RGBD_Miner dataset.

All the experiments were conducted using a PC with 12th Gen Intel(R) Core(TM) i9-12900K, NVIDIA GeForce RTX 3090 with 24G memory and using CUDA 11.3 cuDNN 8.2. The operating system was Windows 10. The study adopted a well-known open source framework, namely the MMDetection framework [21] to train the deep learning models.

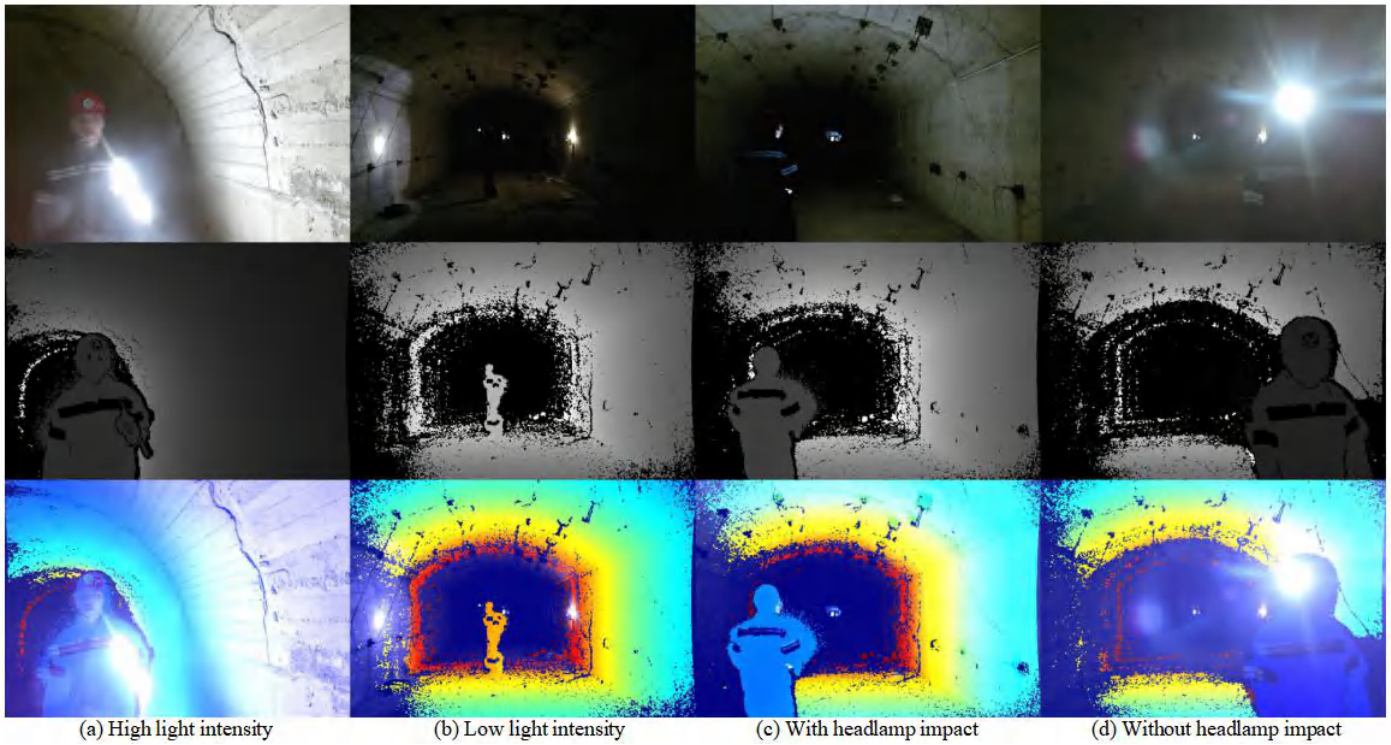


Fig. 5 CUMT_RGBD_Miner Dataset. The first row is the RGB images and the second row is the Depth images. The third row is the images generated by converting the Depth images to pseudo-color images and simply overlaying it with the RGB map, which shows that the RGB image and the Depth image with good alignment accuracy.

B. *Results and Discussion*

Experiments are conducted on the CUMT_RGBD_Miner dataset with YOLOv3, YOLOv3-Depth, methods based on current attention mechanisms, proposed method, etc. The train set and test set have been divided above. Methods are trained on the training set. And the tests are performed on the test set and the performance metrics are shown in Fig. 6 and Table I.

a) Analysis of performance metrics: As shown in Table I, our method is higher than other methods on both AP and AP50, and has approximately same FPS. The viewpoint that fusion of color image information and depth information can improve human detection accuracy in the underground coal mine can be verified from the methods using both information is higher than the method using either information on AP. Our method that uses the improved attention mechanism can more effectively fuse the two types of information can be inferred from the proposed method is higher than the method using current mechanisms on both AP and AP50.As shown in Fig. 6, although the precision of other methods and proposed method is close to 100% when the recall is less than 80%, our method has better performance when the recall is greater than 80%.

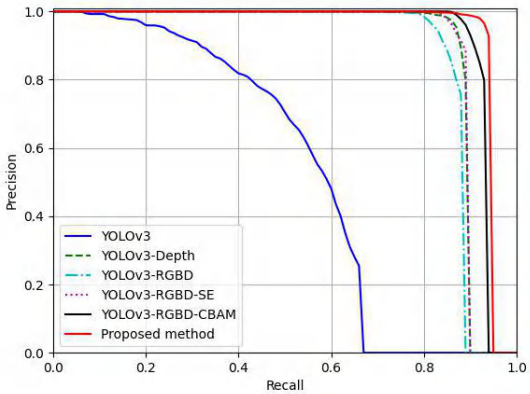
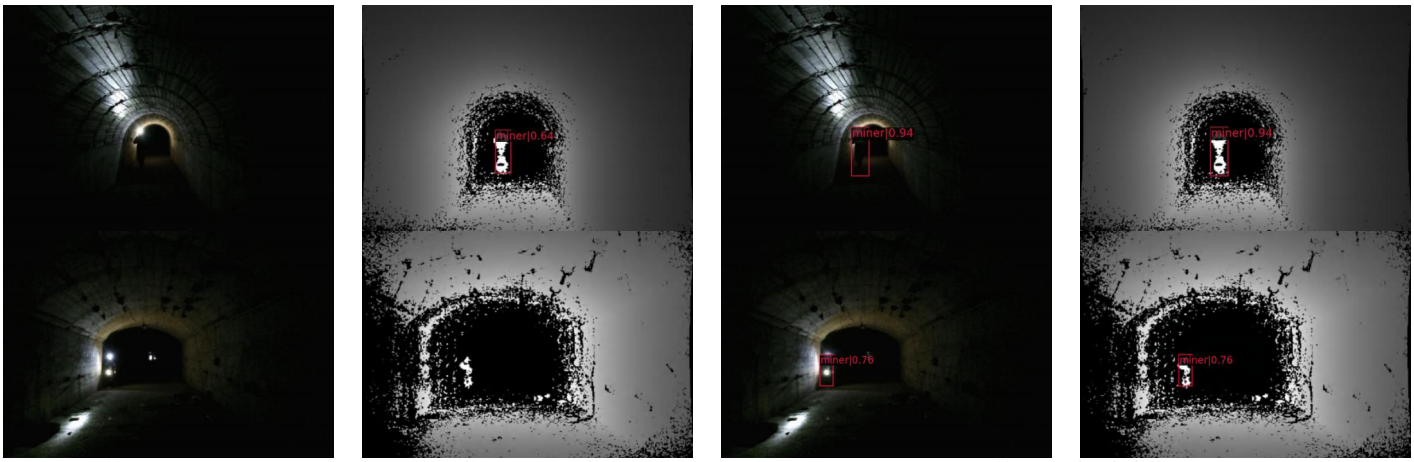


Fig. 6 PR curves for other methods and proposed method.



(a) YOLOv3-RGB (b) YOLOv3-Depth (c) Proposed method

Fig. 7 Detection results of YOLOv3, YOLOv3-Depth and our proposed method.

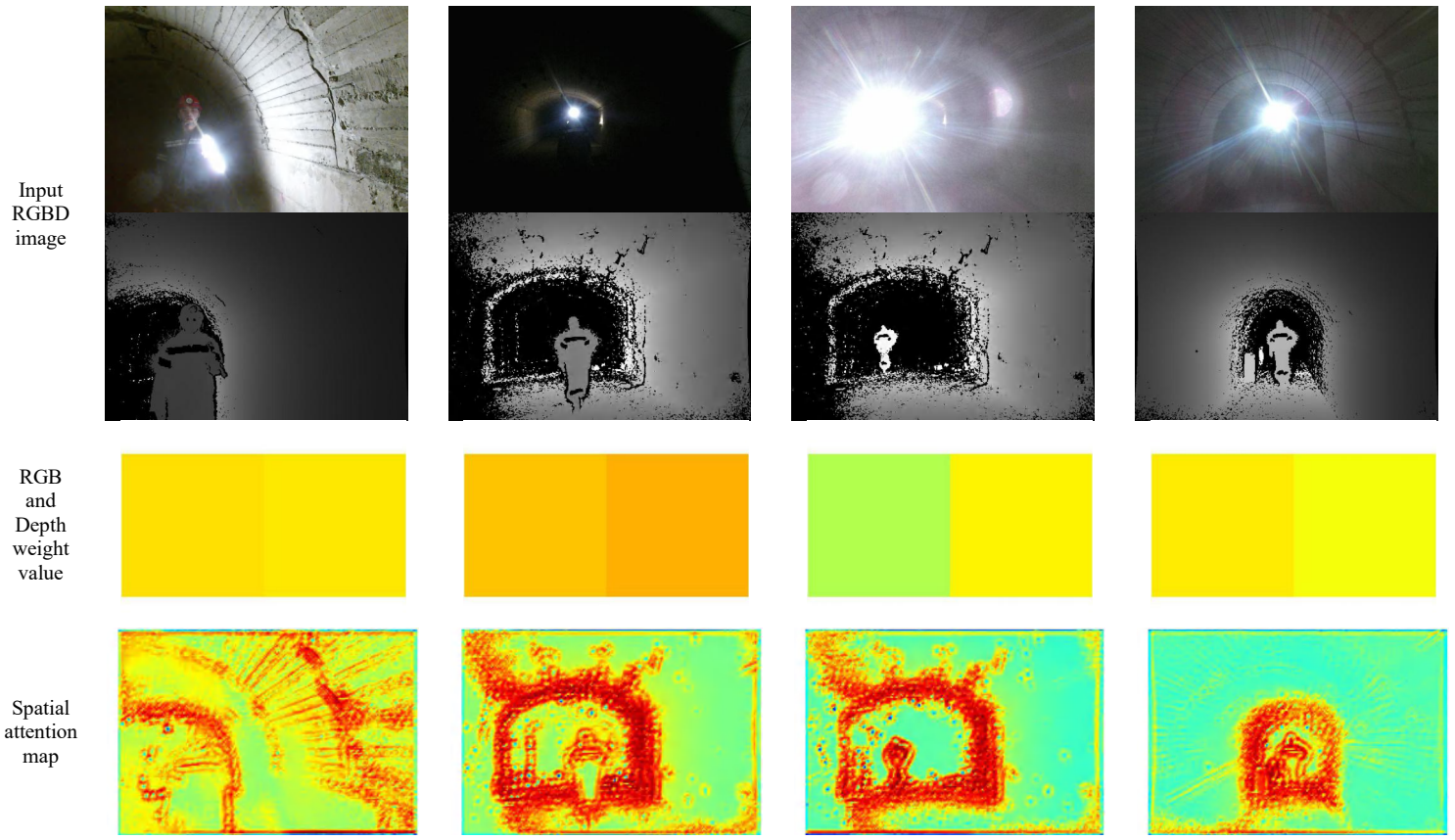


Fig. 8 The visual comparison of weight values and spatial attention maps in the improved CBAM module with RGBD images in different light intensities and inhomogeneities.

b) Comparison of human detection result: Detection results of YOLOv3, YOLOv3-Depth and proposed method on the test set are shown in Fig. 7. Our method has good detection results compared to YOLOv3 which cannot detect human in low light intensity and YOLOv3-Depth which cannot detect human at a long distance. Moreover, our method has higher confidence level when YOLOv3-Depth can also detect human.

c) Validation of the function of improved attention mechanism: We visualized the RGB and Depth weight value

and spatial attention maps obtained in the improved CBAM with RGBD images in different light intensities and inhomogeneities, as shown in Fig. 8. It can be seen that for different light environments, the improved channel attention module has different RGB and Depth weight value, thus achieve the enhancement of useful information and suppression of useless information. And we can see from the spatial attention maps, spatial attention module fuse two types of information to enhance the area where the human is located.

TABLE I
Performance metrics of other methods and proposed method.

Methods	Input size	AP	AP50	FPS
YOLOv3	416x416	21.5	53.9	112
YOLOv3-Depth	416x416	64.9	88.7	118
YOLOv3-RGBD	416x416	67.4	87.2	108
YOLOv3-RGBD-SE	416x416	67.1	88.8	107
YOLOv3-RGBD-CBAM	416x416	70.3	92.5	94
Proposed method	**416x416**	**72.4**	**93.9**	**107**

IV. CONCLUSION

In this paper, a human detection method is proposed with RGBD image within the operating area of electromechanical mining equipment. Our method is improved on YOLOv3. Its framework consists of Pre-Backbone, Post-Backbone and Head. Pre-Backbone extracts the primary RGB and Depth features from the input RGB image and Depth image, selects and preliminarily fuse these two categories of features by enhanced CBAM. Post-Backbone performs deep feature fusion and extraction on the fusion-preliminary features to obtain the final features. Head performs multi-scale prediction on the basis of the final features. With these three components, color and depth information can be efficiently fused for accurate detection of human in underground coal mine. We have conducted experiments on a RGBD dataset. The proposed method show better performance over some classical methods, which demonstrate its superiority on accurate human detection in the environment of low light intensity and uneven light distribution in underground coal mines.

ACKNOWLEDGMENT

This work is supported by the National Science Foundation of China (No. 61903330, No. 62176258 and No. 52174152), the National Key Research and Development Project (2020YFB1314200), the Fundamental Research Funds for the Central Universities (2021YCPY0111), China Postdoctoral Science Foundation (2021M693416) and the Priority Academic Program Development of Jiangsu Higher Education Institution (PAPD).

.

REFERENCES

[1] W. Lan, et al., "Pedestrian Detection Based on YOLO Network Model," 2018 IEEE International Conference on Mechatronics and Automation (ICMA), 2018, pp. 1547-1551.

[2] E. Dong, Y. Zhang and S. Du, "An Automatic Object Detection and Tracking Method Based on Video Surveillance," 2020 IEEE International Conference on Mechatronics and Automation (ICMA), 2020, pp. 1140-1144.

[3] M. Li, Z. Xie, Y. Qin and Y. Mai, "Research on Intrusion Detection Method of Key Metro Areas based on YOLOv3," 2020 International Conference on Sensing, Diagnostics, Prognostics, and Control (SDPC), 2020, pp. 357-360.

[4] L. Cai, et al., "Miner fuzzy detection based on mixture Gaussian model in underground coal mine videos," 2010 3rd International Congress on Image and Signal Processing, 2010, pp. 437-441.

[5] L. Cai and J. Qian, "A Method for Detecting Miners in Underground Coal Mine Videos," 2009 Second International Symposium on Computational Intelligence and Design, 2009, pp. 127-130.

[6] J. Sun, C. Li,"In-pit coal mine personnel uniqueness detection technology based on personnel positioning and face recognition," *Int. J. Min. Sci. Technol.*, vol. 23, no. 3, pp. 357-361, 2013.

[7] P. Jiang, et al., "A Review of Yolo Algorithm Developments," *Procedia Comput. Sci.*, vol. 199, pp. 1066-1073, 2022.

[8] J. Redmon and A. Farhadi, "YOLOv3: An Incremental Improvement," arXiv preprint arXiv:1804.02767, 2018.

[9] X. Yao et al., "Traffic vehicle detection algorithm based on YOLOv3," 2021 International Conference on Intelligent Transportation, Big Data & Smart City (ICITBS), 2021, pp. 47-50.

[10]Y. Li, Q. Wang and R. Liu, "Research on YOLOv3 pedestrian detection algorithm based on channel attention mechanism," 2021 IEEE International Conference on Computer Science, Electronic Information Engineering and Intelligent Control Technology (CEI), 2021, pp. 229-232.

[11]T. Guo, Y. Wei, H. Shao and B. Ma, "Research on Underwater Target Detection Method Based on Improved MSRCP and YOLOv3," 2021 IEEE International Conference on Mechatronics and Automation (ICMA), 2021, pp. 1158-1163.

[12]J. Hu, L. Shen and G. Sun, "Squeeze-and-Excitation Networks," 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2018, pp. 7132-7141.

[13]S. Woo, et al., "Cbam: Convolutional block attention module," Proceedings of the European conference on computer vision (ECCV). 2018, pp. 3-19.

[14]J. Qi, et al., "An improved YOLOv5 model based on visual attention mechanism: Application to recognition of tomato virus disease," Comput. Electron. Agric., vol. 194, 2022.

[15]Y. Li and L. Liu, "YOLO-ResNet: A New Model for Rebar Detection," 2021 4th International Conference on Robotics, Control and Automation Engineering (RCAE), 2021, pp. 128-132.

[16]L. Qu, et al., "RGBD Salient Object Detection via Deep Fusion," in IEEE Transactions on Image Processing, vol. 26, no. 5, pp. 2274-2285, May 2017.

[17]K. He, X. Zhang, S. Ren and J. Sun, "Deep Residual Learning for Image Recognition," 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2016, pp. 770-778.

[18]M. Firman, "RGBD Datasets: Past, Present and Future," 2016 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), 2016, pp. 661-673.

[19]Silberman, Nathan, et al., "Indoor segmentation and support inference from rgbd images." European conference on computer vision. 2012, pp. 746-760.

[20]A. Geiger, P. Lenz, C. Stiller, and R. Urtasun, "Vision meets robotics: The kitti dataset." *Int. J. Rob. Res.*, 2013.

[21]Chen, Kai, et al., "MMDetection: Open mmlab detection toolbox and benchmark," arXiv preprint arXiv:1906.07155, 2019.