ISSN No: 2348-4845
International Journal & Magazine of Engineering, Technology, Management and Research
A Peer Reviewed Open Access International Journal

Objective detection using convolutional neural network based on the vanishing point for the high resolution images

K Ramanjaneyulu
Research Scholar, ECE Department
JNTUCEK, Kakinada
ramanjaneyulu.k@qisit.edu.in

K Veera Swamy
Professor, ECE Department
Vasavi College of Engineering,
Hyderabad, Telangana
k.veeraswamy@staff.vce.edu.in

CH. Srinivasa Rao
Professor, ECE Department
JNTUKCEV, Vijayanagaram,
Andhra Pradesh
ch_rao@rediffmail.com.

Abstract **–** Object detection is the most important role for the applications of monitoring security system, medical, and entertainment etc. Generally, images are two types of resolutions. One is high- resolution and low-resolution. In the high-resolution process down-sampling process is the key role for the extracting the features from the input images and reduce the computational complexity but it does not cover the missed and distant objects. To overcome this problem, the proposed method has been successfully implemented with the help of vanishing points frameworks. The features are extracted from various deep convolutional neural networks with the help of activation functions. It has been implemented with the real time images from the railway surveillance monitoring system. This proposed method provides better results in terms of performance measures over existing methods.

Keywords: object detection, vanishing points, convolutional neural network

I.INTRODUCTION

Video surveillance plays an important role in operation safety of railway, especially for intruding foreign object recognition. Object detection is an essential step for railway intrusion detection and a basic step for further analysis, such as action recognition. Traditional object detection methods mainly rely on background subtraction [1] that separates moving objects from static background. But railway scene is captured from outside, it contains several challenges for object detection, such as light changing, nighttime and camera shaking. Deep learning, especially deep CNN, achieves excellent results on large scale image classification task [2-3], object detection [4-6] and image segmentation [7]. The data-driven supervised methods learn features form dataset that are more discriminative and robust than the hand-crafted features that traditional methods use. So, deep CNN is used to detect object in railway scene that is robust to light changing and camera shaking [8].

However, CNN computation is expensive and the high-resolution input before feeding into CNN commonlyneeds to be scaled down that often leads distant small object missed detection, which causes serious railway safety problems. Meanwhile, the purpose of camera installation for railway monitoring system is to make monitoring scene as large as possible, there exists a vanishing point (VP) inside the scene, which is that paralleled lines in 3D real

Cite this article as: K Ramanjaneyulu, K Veera Swamy & CH. Srinivasa Rao, "Objective detection using convolutional neural network based on the vanishing point for the high resolution images", International Journal & Magazine of Engineering, Technology, Management and Research, Volume 6, Issue 4, 2019, Page 182-188.

ISSN No: 2348-4845
International Journal & Magazine of Engineering, Technology, Management and Research
A Peer Reviewed Open Access International Journal

world intersect on 2D image plane and provides scene depth information. Small objects often occur in the distant region around VP, shown in Fig.I Besides scaled original input, we also extract the distant region as an auxiliary and additional input for object detection to preserve small object details. VP also appears in highway scene

Fig. 1. A railway scene (1920*1080) contains intruding Foreign objects (red rectangle) and a vanishing point (green circle).

from on-board camera that is used for autonomous driving navigation. Traditional VP detection methods [9-12] mainly consist of hand-crafted feature extraction and voting or clustering steps, which are sensitive to disturbance and time-consuming. However, VP detection in railway scene is a complex problem that contains not only complex light conditions and noisy scenes, but also the various scene content unlike highway scene that is slightly fixed. Because of powerful features extraction of deep CNN, authors [13-14] used CNN to predict VP. Chang C.K. et al. [13] formulated VP detection as a classification problem. They divided the image into N*N cells and numbered each cell.After that, a VP was mapped into a cell number as its class label according to the coordinates. Y. Shuai et al. [14] used AlexNet structure that last fully connected layer had two neurons to regress VP coordinates directly that were normalized to the range of 0-1. But directly predicting VP with a regression network is non-trivial, especially on noisy scenes.

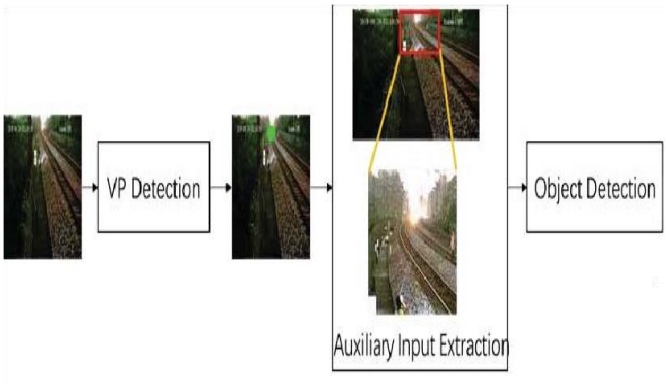

Fig. 2. The proposed framework that extracts auxiliary input around vanishing point to preserve small object details

Classification model with N*N neurons in last layer is slightly complex and may cause class imbalance problem because of some classes without corresponding samples. So, a two-branch network that classifies horizontal and vertical coordinates respectively could alleviate this problem.

In this paper, based on VP detection, we proposed a framework that extracted distant region in railway scene as an auxiliary input to improve distant small object detection, as shown in Fig. 2. We improved one-branch VP classification model by replacing last layer with two fully connected layers that classified horizontal and horizontal coordinates respectively. The rest of the paper is organized as follows. The Section Ċ introduces the proposed framework, including VP detection and the auxiliary input extraction. The Section ċ shows the experimental results and Section Č draws conclusions.

ISSN No: 2348-4845
International Journal & Magazine of Engineering, Technology, Management and Research
A Peer Reviewed Open Access International Journal

II. THE PROPOSED FRAMEWORK

A. Vanishing point detection

We formulate VP detection as a CNN-based classification problem. First, we divide a T*T image into N*N cells and each cell size is S*S, i.e. S=T/N. Then we number each cell from the x coordinate (horizontal direction) and the y coordinate (vertical direction) instead of each cell being regarded as a different class in [13], as shown in Fig. 3. The cell numbers (x, y) VP located in are labeled classes, (x, y)□□{0,1,2..., N -1} . Finally, we use a two-branch CNN model to predict the x-coordinate and y-coordinate classes (x1, y1).Our two-branch model has two separate branches in the final layer totally with 2*N neurons,which is slightly simple than one-branch model with N*N neurons, especially in case of lager W and smaller S.Final predicted VP is a small rectangle with top left point ($x1*S, y1*S$) and bottom right point ($x1*S+S, y1*S+S$) .

Fig. 3. Discretize VP coordinates into classes from the x coordinate and the y coordinate. The vanishing point (redbox) classes are (7,1). T=300. N=15.

The two-branch model is based on AlexNet, as well as [13], except for last layer, as shown in Fig. 4. It mainly composes of five convolutional layers (Conv) and four fully connected layers (FC) that followed by rectified linear units (ReLU). Three Max polling layers (MaxPool) re add after Conv ķ

$$loss^{(i)} = -\sum_{c=0}^{N} x_c^{(i)} \log(px_c^{(i)}) - \sum_{c=0}^{N} y_c^{(i)} \log(py_c^{(i)})$$

Conv к Conv Ļ layers. The wo-branch model can also be regard as a multi-task earning problem, which predicts x coordinate and y oordinate simultaneously. So, the loss is the sum of the two tasks.

N is the number of classes. (x_c, y_c) are binary indicators (0 or 1) if class label c is the correct classification for sample i. (px_c, py_c) are predicted probabilities of sample i and class c.

B. *The auxiliary input extraction*

Extracting distant region around VP of a high-resolution railway scene before scaling down, as an auxiliary input, is important to object detection, especially for small objects, which often occurs in distant region. InVP detection, we often scale a *W*H* image into *T*T*. So, we should map the predicted VP position to original imageaccording to:

$$lx = x1*S*\frac{W}{T}, ly = y1*S*\frac{H}{T}$$

$$rx = (x1*S+S)*\frac{W}{T}, ry = (y1*S+S)*\frac{H}{T}$$

The mapped VP position is a rectangle with top left point (lx,ly) and bottom right point (rx,ry) . The final Auxiliary input is a region $(lx \square len, ly \square len, rx \square len, ry \square len)$ outside of the VP rectangle with a fixed length len. This region acts as an auxiliary input feeding into object detection network.

ISSN No: 2348-4845
International Journal & Magazine of Engineering, Technology, Management and Research
A Peer Reviewed Open Access International Journal

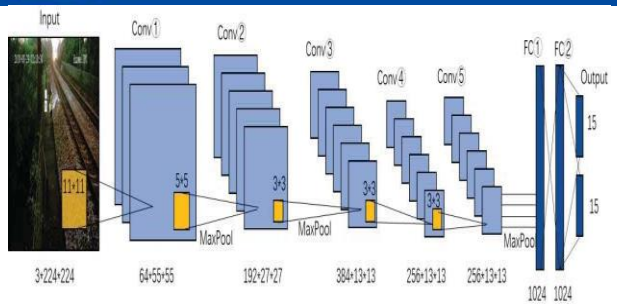

Fig. 4. The two-branch vanishing point detection network structure (for N=15).

III. EXPERIMENT

We train our model on larger Road Dataset (RoD) and then transfer it to smaller Railway Dataset (RaD). After VP detection, we extract the auxiliary input and then feed it into a trained object detection network SSD to boost small object detection.

A. Dataset

The dataset is composed of RoD and RaD. RoD are from [13], consisting of 23 route scenes. Because adjacent number frames in each scene is similar, we sample images from each scene and put them together to make dataset various. We collected 28,660 frames. We collected RaD from website and actual railway scenes, 527 images from Google Images and Baidu Images that each image scene is quite different, and 268 images from different views and times of 17 railway scenes. We make labels based on image size (300*300) and resize them to 224*224 before feeding them into the model. And we split them 4:1 for training and validation.

B. Training configuration

We trained the model with Adam optimizer, 50 epochs, early stopping when accuracy has not been improved lasting for 5 epochs, two dropout layers following FC I and FC 2 and learning rate reducing with times 0.1 when validation loss has not been improved lasting for 1 epoch. The remaining configurations are different between RoD and RaD, shown in table I

Table I. Training configurations for Road Dataset (RoD) and Railway Dataset (RaD).

Configuration	RoD	RaD
Initialization	Kaiming Initialization [15]	Pre-trained from RoD
Learning Rate	0.0003	0.001
Batch Size	128	4
Dropout Rate	0.5	0.7

Based on predicted labels $(x1, y1)$ and ground truth (x, y), we defined the average error distance (AED) to evaluate the performance. Where NUM is sample numbers.

$$AED = \frac{1}{NUM}\sum_{NUM}\sqrt{(x*S - x1*S)^2 + (y*S - y1*S)^2}$$

C. VP detection

We first trained our model on RoD and compared with one-branch model [13] and regression model [14] based on the same AlexNet structure except for last layer that have N*N neurons and 2 neurons respectively. We also evaluated performances on different N=15, 20, 30, 60 for two classification models. The AED, parameter numbers and memory are shown in Table II-IV. We find that N=20 has smallest AED for classification models. Cell size S and classification accuracy affect AED. VP prediction based on N=15 is not precise, but too larger N leads classification accuracy decreasing because small S has not enough VP information. The feature maps of the ConvL_5 of three samples are shown in Fig. 5, which are summed along the channel direction proving that convolutional filters are sensitive to VP region. That AED of one-branch model

ISSN No: 2348-4845
International Journal & Magazine of Engineering, Technology, Management and Research
A Peer Reviewed Open Access International Journal

decreasing in N=60 compared with N=30 is caused by the smaller degree of accuracy decreasing than cell size, so the product is low than N=30. Meanwhile, our two-branch model has smaller AED, parameters and memory usage. One-branch model has more classes, which are the N/2 times more than ours, but a big part of them have no corresponding samples, which leads class imbalanced problem. What's more, two-branch model is slightly simple, so it achieves smaller AED. Some visual results are shown in Fig. 6..

Table II. Average Error distance (AED)of regressionnetwork, one-branch and two-branch model.

Model	N			
	15	20	30	60
One-Branch	9.98	8.81	9.73	9.24
Two-Branch	8.81	8.09	8.37	8.71
Regression	23.17			

For regression model, despite of its small parameters and memory usage, it gets larger AED with the same network structure, it maybe needs deeper network, after all, directly predicting VP coordinates is non-trivial

Table III. Parameters of regression network, one-branch and two-branch model.

Model	N			
	15	20	30	60
One-Branch	13,188,129	13,367,504	13,880,004	16,647,504
Two-Branch	12,988,254	12,988,504	13,019,004	13,080,504
Regression	12,959,544			

Table IV. Memory usage (MB) of regression network, one-branch and two-branch model.

Model	N			
	15	20	30	60
One-Branch	59.14	59.82	61.78	72.36
Two-Branch	58.37	58.41	58.49	59.73
Regression	58.26			

Fig. 5. Feature maps of the ConvL5

Fig. 6. Visual results between one-branch model (blue box) and two-branch model (red box) on Road Dataset (RoD). Green box is ground truth

we transferred two-branch model to RaD. The validation AED for website images and railway images are 14.77 and 36.97.Because website images are quite different between training & validation, so the error is large. However, actual images come from several fixed scenes, their AED is low and enough for auxiliary input extraction. Some visual results are shown in Fig. 7.

ISSN No: 2348-4845
International Journal & Magazine of Engineering, Technology, Management and Research
A Peer Reviewed Open Access International Journal

Fig. 7. Visual results of two-branch model (red box) on Railway Dataset (RaD).Green box is ground truth.

Fig. 8. Auxiliary input extraction (green rectangle) for distant small object detection. Red rectangle is mapped VPregion. Object inside white ellipse is small and occurs in distant region around vanishing point. The object is missedin original input but detected in auxiliary input.

D.The auxiliary input extraction

After VP prediction, we mapped predicted VP labels to original high-resolution image according to section II-B and extracted a fixed region (green rectangle) centered on VP region (red rectangle) as the auxiliary input, such as 400*400, shown in Fig. 8. In this experiment, we feed the auxiliary input to a trained object detection network SSD [5] with input size 300*300. The distant small object in white ellipse is missed in original input but detected in auxiliary input.

IV.CONCLUSION

We proposed a framework leverage VP information to produce auxiliary input for distant small object detection. The auxiliary input around VP makes object detection network do not miss distant objects, which is important to railway intrusion detection. Our two-branch model achieves lower VP detection error than one-branch model and regression model. Our proposed framework is positive to the robustness of railway surveillance system.

ACKNOWLEDGMENT

Authors would like to express gratitude to Dr.S. Srinivasa Kumar, Former -Vice-chancellor, JNTUA, Anantapur, India for his precious suggestions.

REFERENCES

[1] T. Bouwmans, "Traditional and recent approaches in background modeling for foreground detection: An overview," *Computerscience review*, vol. 11, pp. 31–66, 2014.

[2] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classifcation with deep convolutional neural networks," *in Advances in neural information processing systems*, 2012, pp. 1097–1105.

[3] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," *in Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 770–778.

[4] S. Ren, K. He, R. Girshick, and J. Sun, "Faster r-cnn: Towards real-time object detection with region proposal networks," *in Advances in neural information processing systems*, 2015, pp. 91–99.

[5] W. Liu, D. Anguelov, D. Erhan, C. Szegedy, S. Reed, C.-Y. Fu, and A. C. Berg, "Ssd: Single shot multibox detector," *in European conference on computer vision. Springer*, 2016, pp. 21–37.

[6] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, "You only look once: Unifed, real-time object detection," *in Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 779–788.

ISSN No: 2348-4845
International Journal & Magazine of Engineering, Technology, Management and Research
A Peer Reviewed Open Access International Journal

[7] J. Long, E. Shelhamer, and T. Darrell, "Fully convolutional networks for semantic segmentation," *in Proceedings of the IEEE conference on computer vision and pattern recognition*, 2015, pp. 3431–3440.

[8] W. Yang, Y. Zujun, Z. Liqiang, and G. Baoqing, "Fast feature extraction algorithm for high-speed railway clearance intruding objects based on cnn," *Chinese Journal of Scientifc Instrument*, vol. 38, no. 5, pp. 1267–1275, 2017.

[9] J. H. Yoo, S.-W. Lee, S.-K. Park, and D. H. Kim, "A robust lane detection method based on vanishing point estimation using the relevance of line segments," *IEEE Transactions on Intelligent Transportation Systems*, vol. 18, no. 12, pp. 3254–3266, 2017.

[10] W. Yang, B. Fang, and Y. Y. Tang, "Fast and accurate vanishing point detection and its application in inverse perspectivemapping of structured road," *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, vol. 48, no. 5, pp. 755–766, 2018.

[11] H. Kong, J.-Y. Audibert, and J. Ponce, "General road detection from a single image," *IEEE Transactions on Image Processing*, vol. 19, no. 8, pp. 2211 – 2220, August 2010.

[12] Y. Wang, E. K. Teoh, and D. Shen, "Lane detection and tracking using b-snake," *Image and Vision computing*, vol. 22, no. 4, pp. 269–280, 2004.

[13] C.-K.Chang, J. Zhao, and L. Itti, "Deepvp: Deep learning for vanishing point detection on 1 million street view images," *in 2018 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, 2018, pp. 1–8.

[14] Y. Shuai, Y. Tiantian, Y. Guodong, and L. Zize, "Regression convolutional network for vanishing point detection," *in 201732nd Youth Academic Annual Conference of Chinese Association of Automation (YAC). IEEE*, 2017, pp. 634–638.

[15] K. He, X. Zhang, S. Ren, and S. Jian, "Delving deep into rectifers: Surpassing human-level performance on imagenet classifcation," *in Proceedings of the IEEE international conference on computervision*, 2015. pp.