

Racing Bib Number Localization Based on Region Convolutional Neural Networks

Noppakun Boonsim and Saranya Kanjaruek

Faculty of Applied science and Engineering, Khon Kaen university, Thailand

Abstract. This work aims to apply region convolutional neural network technique for detecting racing bib numbers on marathon images. Basically, marathon images are captured in the outdoor environment. The images are often ruined because of environment variations such as illumination, occlusion, and complex background. Transfer learning technique is applied for detecting a racing bib number on marathon images. Transfer learning algorithm firstly loads pre-trained model that was already trained on Cifar-10 image dataset. After that, racing bib number images (training images) are trained together with the pre-trained model into two classes: racing bib number class and non-racing bib number class. Finally, the trained model from the previous step is used to detect a racing bib number in marathon images. The results were reported to overtake previous research and F-measure was at 90 % on 400 marathon images.

Keywords: racing bib number, localization, deep learning, region convolutional neural network.

1. Introduction

At the present time, there are many marathon events all around the world. Participants are pinned by unique number tag on their shirt, which can be called Racing Bib Number (RBN) and is used to identify each runner. For security and management reasons, organizers ensure that participants are identified by RBN by collecting images of a runner. With a lot of participants and images in a marathon event, it is challenging to identify participants manually. To solve this problem, this work applied region convolutional neural network techniques to locate RBN in runners' images which can be used to accurately recognize RBN.

The first RBN recognition process is RBN detection, a process to locate the RBN position within an image. In RBN segmentation process, RBN's are separated one by one in order to prepare RBN for recognition in the next process. RBN recognition process aims to recognize each RBN. For improving accuracy in RBN recognition, this work addresses the RBN detection, the first process in the pipeline and an important process of the RBN system, because the performance of RBN recognition depends on the accuracy of RBN detection.

Many proposed techniques to detect text in natural scene images have been developed to date [1]. In marathon perspective, there are three studies dedicated to detecting RBN. Stroke width transform technique extracted characters from an image, which was proposed by Ami et al. [2]. RBN candidates grouped characters with similar stroke width into a region. The Stroke Width Transform technique uses runner's body to verify the real RBN for detecting runner's body. Stroke Width Transform algorithm first detects runner's face and then locates runner's body. The detection accuracy of Stroke Width Transform technique is at 45% (f-measure). To improve the accuracy, Roy et al. [3] proposed a multi-modal technique. First, face and skin color are used to detect runner's body in natural and marathon images. Following, wavelet and color features are extracted from the given runner's body. K-mean clustering technique is used to finally locate the true RBN location. Binarization technique [4] and Tesseract OCR technique [5] are employed to recognize RBN. This study was reported detection accuracy at 64 % which tested on 212 marathon images. Shivakumara et al.

⁺ Corresponding author. Tel.: +66817296239; fax: +6642415699.
E-mail address: kanjaruek@kku.ac.th.

[6] presented a combining technique which is also called multi-modal technique. The technique starts with torso detection. Texts within torso are considered as candidates or potential texts. Shivakumara et al. extracted text features in RBN by using Fourier feature, Polar feature, and Zernike feature detectors and classified RBN by using Support Vector Machine (SVM) technique. Result texts from previous steps are classified one more time by using Histogram of Gradient (HOG) feature detector in order to increase RBN detection accuracy. Moreover, the remaining unidentified texts are extracted with HOG features and are classified by SVM. The detection accuracy of the multi-modal technique was reported at 66%.

This work explores how deep learning technique improves detection accuracy of RBN position. Deep learning technique (learning from scratch, transfer learning, and feature extraction) solves a problem by learning from experiences (training dataset) and shows excellent performance and outperforms machine learning techniques on object detection segmentation, recognition, and image classification [7]. Learning method from scratch method highlights training neural network from the beginning. Transfer learning method (pre-trained model) defines a trained new model to solve a new problem from the previous model. Transfer learning method is faster and easier than training from scratch method and it reduces the amount of training data from one thousand to one hundred images (training a new network using a million images) because it utilizes pre-trained model. Feature extraction method employs or acquires features from trained networks as input to machine learning technique (decision tree, nearest neighbor, and support vector machine).

Transfer learning technique is selected to solve the problem of detecting RBN on marathon images because it is suitable for the limited datasets and requires a small amount of time for training the network. Region convolutional neural network is designated to recognize a specific range of RBN image. The purposed method trains detection model by training images of RBN images (positive images) and other images (negative images). Furthermore, the trained model classifies RBN images from other images (graphics, symbols, and brand names).

The remaining sections are organized as follows: the proposed method and details are given in section 2. Section 3 illustrates the experimental data, tools, results, and discussions. The conclusion of this research is presented in section 4.

2. The Proposed Method

Machine learning extracts patterns from raw data in order to acquire their own knowledge. The more given data is represented; the better performance of machine learning algorithm is presented. Thus, many artificial intelligence tasks can be tackled by creating the right set of features to extract for that task and giving these features to a machine learning algorithm. However, it is difficult to select what features should be extracted from raw data for many tasks.

Deep learning uses multi-layered perceptron, which allows a computer to learn a multi-step computer program. A variety of different methods has been proposed for image detection tasks. The deep learning technique in image detection is able to implement both local feature and global feature. A local feature means focusing on the specific region of interest, while a global feature emphasizes on the entire whole image. Considerable attention must be paid when the deep learning technique is employed to detect RBN position within an image. This method advances the previous methods by using less computation time and processing a limited dataset.

This present paper purposes, RBN Finder, the usage of region deep learning method (Region Convolutional Neural Network (R-CNN) [8], as presented in Figure 1) to detect RBN on marathon shirt on natural scene. RBN consists of a region with a rectangular shape and a group of characters.

RBN Finder consists of three steps. First, pre-trained model (negative instances) using the CIFAR-10 dataset [9] is loaded into memory. The CIFAR-10 dataset has 50,000 training images with 10 image classes, and each class has 5,000 images.

Step two is using the output from the first step, a pre-trained model. RBN Finder applied LeNet [10] to a pre-trained network. LeNet encloses 15 network layers (one layer of Input layer, fully connected layer, Softmax, Classification output layer and three layers of Convolution layer, three layers of Relu layer and three

layers of Max pooling layer). The fully connected layer is fine-tuned network that can perform classification into two classes: RBN class and non-RBN class. These are instructed as a new model for detecting RBN position. RBN class includes varieties of RBN images from a marathon event (positives instances as illustrated in Figure 1).

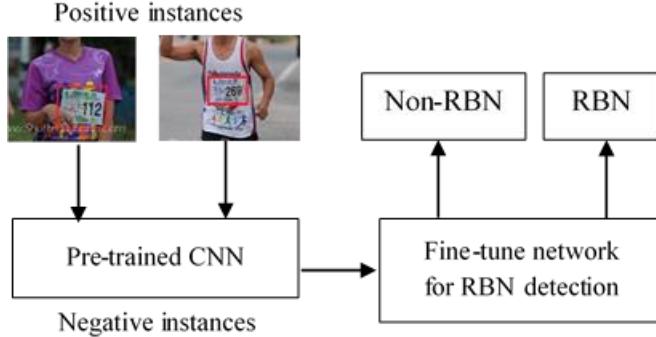


Fig. 1: RBN detection system overview

The training RBN images include normal RBN images, rotated RBN images, occlusion RBN images, sheared RBN images, and low contrast RBN images. The non-RBN class refers to the CIFAR-10 dataset.

The pre-trained model applies Stochastic Gradient Descent with Momentum (SGDM) with an initial learning rate of 0.001 for adjustment purpose. The initial learning rate is reduced every 8 epochs (1 epoch is defined as one complete pass through the entire training dataset) during the training process. The pre-trained algorithm processes 40 epochs as default.

Final step, the fine-tuned network from the second step is trained with the same algorithm from the first step with the initial learning rate of the pre-trained network. The training algorithm processes for 100 epochs and illustrates the satisfying result on the training dataset.

3. Experimental data, tools, results, and discussions

The experimental result of proposed work has the capacity of 2.24 second with 400 images. Precision and recall are 0.93 and 0.87 respectively. F-measurement is 0.90 as presented in table 1.

Table 1: The performance of the proposed approach and existing approach on bib number detection

Works	Images	Technique	Time(s)	P	R	F
Ami et al.	217	SWT + face	N/A	0.53	0.40	0.45
Roy et al.	212	Multi-modal	N/A	0.76	0.55	0.64
Shivakumara et al.	212	Multi-modal	N/A	0.60	0.74	0.66
RBN Finder	400	RCNN	2.24	0.93	0.87	0.90

Table 1 illustrates a comparison of RBN Finder with other works by evaluating image data set, a technique used, computational time, and detection rate (precision, recall, and f-measure). Ami et al. [2] focuses on the stroke width transform to find text with related stroke width within an RBN image. Ami et al. located runner's upper body using face detection. The Ami et al. work reported the detection accuracy (f-measure) at 45% on 217 images with details of computational time.

The multi-modal technique [3], a number of steps to detect the real RBN, improved the accuracy to 64%, employed on 212 images, but there is no time of computation described. Disadvantage of this technique, the color based method is utilized which is very sensitive to lighting conditions. On the other hand, many features and classifiers were precisely used for detecting RBN by the multi-modal technique [6]. The detection accuracy was reported at 66%, tested on 212 images. Whereas RBN Finder employs deep learning technique (Region Convolutional Neural Networks) to detection RBN position in marathon images. RBN Finder experiments were tested on 400 images from a Benchmark for Enhancing the Standard of Thai language processing (BEST) [11] with 2.24 seconds of computational time per image. The detection accuracy of RBN Finder was testified at 90%. The advantages of RBN Finder method are enable to detect

the partial occlusion RBN, as shown in Figure 2(b), to detect the rotated RBN, Figure 2(c), and to detect the multiple RBN tags or runners which can be seen in Figure 2(d).

However, RBN Finder has a small number of problems with marathon images that having object similar to RBN (Figure 3(a)-(c)) and text similar to RBN tag (Figure 3(d)). Figure 2(a) - Figure 2(d) displayed the virtuous results of RBN Finder. The example of images that contain object similar to RBN tag in figure 3(a)-(c) and the example of detected other text 3(d) which is the false alarm.



Fig. 2: The example of achievement detection

4. Conclusion

This research presented a method to locate RBN position in marathon images, which is the important step of RBN recognition system. The RBN Finder method is based on deep learning technique. The challenge with RBN Finder method is producing false alarms when implemented on an extremely complex background, e.g. many texts and containing object similar to RBN tag. In addition, RBN Finder method, deep learning technique, is need a lot of training images to train model. Future work will find a new method to extract RBN appearances or features will improve the accuracy of the process. Future work will attempt to solve these problems, which will improve the performance by using the CIFAR-100 dataset.





3(c) detect license plate (similar shape)



3(d) detect other text (similar texture)

Fig. 3: The example of Fail detection

5. References

- [1] Q. Ye and D. Doermann, "Text detection and recognition in imagery: A survey," *IEEE Trans. Pattern Analysis and Machine Intelligence.*, vol. 37, no. 7, pp. 1480–1500, 2015.
- [2] I. Ben-Ami, T. Basha, and S. Avidan, "Racing Bib Numbers Recognition," in *BMVC*, 2012, pp. 1–10.
- [3] S. Roy, P. Shivakumara, P. Mondal, R. Raghavendra, U. Pal, and T. Lu, "A New Multi-modal Technique for Bib Number/Text Detection in Natural Images," in *Pacific Rim Conference on Multimedia*, 2015, pp. 483–494.
- [4] N. R. Howe, "Document binarization with automatic parameter tuning," *Int. J. Doc. Anal. recognition. IJDAR*, vol. 16, no. 3, pp. 247–258, 2013.
- [5] R. Smith, "An overview of the Tesseract OCR engine," in *Document Analysis and Recognition, 2007. ICDAR 2007. Ninth International Conference on*, 2007, vol. 2, pp. 629–633.
- [6] P. Shivakumara, R. Raghavendra, L. Qin, K. B. Raja, T. Lu, and U. Pal, "A new multi-modal approach to bib number/text detection and recognition in Marathon images," *Pattern Recognition.*, vol. 61, pp. 479–491, 2017.
- [7] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," in *Advances in neural information processing systems*, 2012, pp. 1097–1105.
- [8] R. Girshick, J. Donahue, T. Darrell, and J. Malik, "Rich feature hierarchies for accurate object detection and semantic segmentation," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2014, pp. 580–587.
- [9] A. Krizhevsky and G. Hinton, "Learning multiple layers of features from tiny images," 2009.
- [10] Y. LeCun, L. Bottou, Y. Bengio, and P. Haffner, "Gradient-based learning applied to document recognition," *Proc. IEEE*, vol. 86, no. 11, pp. 2278–2324, 1998.
- [11] Sinthupinyo, W. (2016). BEST. [online] Thailand.nectec.or.th. Available at: <https://thailang.nectec.or.th/best> [Accessed 10 Mar. 2018].